

UNIVERSITY  
OF MICHIGAN

APR 2 1953

PERIODICAL  
READING ROOM

# Review of Educational Research

VOL. XXIII, No. 1

FEBRUARY 1953

*desk*

EDUCATIONAL AND PSYCHOLOGICAL TESTING

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

A Department of the

NATIONAL EDUCATION ASSOCIATION OF THE UNITED STATES

1201 Sixteenth St., N.W., Washington 6, D. C.

## AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

THIS ASSOCIATION is composed of persons engaged in technical research in education, including directors of research in school systems, instructors in educational institutions, and research workers connected with private educational agencies.

### Executive Committee, February 1952—February 1953

**President:** ARVIL S. BARR, Professor of Education, University of Wisconsin, Madison, Wisconsin.

**Vicepresident:** GUY T. BUSWELL, Department of Education, University of California, Berkeley, California.

**Secretary-Treasurer:** FRANK W. HUBBARD, Director, Research Division, National Education Association, Washington 6, D. C.

**Immediate Past President:** PAUL R. MORT, Professor of Education, Teachers College, Columbia University, New York, New York.

**Chairman and Editor of the Review:** FRANCIS G. CORNELL, Professor of Education, University of Illinois, Urbana, Illinois.

**Chairman and Editor of the Newsletter:** GERALD V. LANNHOLM, Project Director, Educational Testing Service, Princeton, New Jersey.

### Editorial Board of the Review

The chairman and editor,\* the president and the secretary-treasurer.

ROBERT J. HAVICHURST, Professor of Education, University of Chicago, Chicago, Illinois.

GORDON N. MACKENZIE, Professor of Education, Teachers College, Columbia University, New York 27, New York.

\*Assistant Editor: DARRELL J. INABNIT, III, University of Illinois, Urbana, Illinois.

*Applications* for membership should be sent to the secretary-treasurer. Upon approval by a committee of the Association, persons applying will be invited to become members.

*Subscriptions* to the REVIEW should be sent to the secretary-treasurer (note address above).

*Orders* for one or more publications, accompanied by funds in payment, should be sent to the American Educational Research Association, 1201 Sixteenth St., N. W., Washington 6, D. C. For a list of titles see the back inside cover page.

---

Active and associate members of the Association pay dues of \$8 annually. Of this amount \$5 is for subscription to the REVIEW. The REVIEW is published in February, April, June, October, and December. Beginning with the February 1949 issue single copies are priced at \$1.50.

---

Entered as second-class matter April 10, 1931, at the post office at Washington, D. C., under the Act of August 24, 1912.

General Lib.  
Direct

# REVIEW OF EDUCATIONAL RESEARCH

Official Publication of the American Educational Research Association.

Contents are listed in the Education Index.

Copyright 1953

By National Education Association of the United States, Washington, D. C.

L  
11  
R454

Vol. XXIII, No. 1

February 1953

## Educational and Psychological Testing

Reviews of the literature for the three-year period since the issuance of Vol. XX, No. 1, February 1950.

### TABLE OF CONTENTS

Chapter	Page
Foreword .....	4
I. Testing and the Use of Test Results .....	5
FREDERICK B. DAVIS, <i>Hunter College, New York, New York</i>	
II. Development and Applications of Test of General Mental Ability .....	11
JULIAN C. STANLEY, <i>George Peabody College for Teachers, Nashville, Tennessee</i>	
III. Development and Applications of Tests of Special Aptitude ...	33
WILLIAM G. MOLLENKOPF, <i>Educational Testing Service, Princeton, New Jersey</i>	
IV. Development and Applications of Nonprojective Tests of Personality and Interest .....	56
DAVID V. TIEDEMAN, <i>Harvard University, Cambridge, Massachusetts</i>	
KENNETH M. WILSON, <i>Harvard University, Cambridge, Massachusetts</i>	
V. Development and Applications of Projective Tests of Personality .....	70
JOHN W. M. ROTHNEY, <i>University of Wisconsin, Madison, Wisconsin</i>	
ROBERT A. HEIMANN, <i>Arizona State College, Tempe, Arizona</i>	

<i>Chapter</i>	<i>Page</i>
VI. Development and Applications of Tests of Educational Achievement in Schools and Colleges.....	85
ERIC F. GARDNER, <i>Syracuse University, Syracuse, New York</i>	
VII. Development and Applications of Tests of Educational Achievement Outside the Schools.....	102
JOHN T. DAILEY, <i>Department of the Navy, Washington, D. C.</i>	
Index .....	110

This issue was prepared by the Committee on  
Educational and Psychological Testing

FREDERICK B. DAVIS, *Chairman*, Hunter College, New York, New York

JOHN T. DAILEY, Department of the Navy, Washington, D. C.

ERIC F. GARDNER, Syracuse University, Syracuse, New York

JULIAN C. STANLEY, George Peabody College for Teachers, Nashville,  
Tennessee

DAVID V. TIEDEMAN, Harvard University, Cambridge, Massachusetts

with the assistance of

ROBERT A. HEIMANN, Arizona State College, Tempe, Arizona

WILLIAM G. MOLLENKOPF, Educational Testing Service, Princeton, New  
Jersey

JOHN W. M. ROTHNEY, University of Wisconsin, Madison, Wisconsin

KENNETH M. WILSON, Harvard University, Cambridge, Massachusetts

## FOREWORD

**R**ESearch studies dealing with educational and psychological testing continue to appear in ever-increasing volume. In the preparation of this issue, considerable selectivity was exercised in order to keep within space limitations. Efforts were made to avoid references dealing mainly with methods of research and experimentation; the last issue of the REVIEW covering this area appeared in December 1951 as Volume XXI, No. 5.

During the three years since the appearance of the last issue of the REVIEW on educational and psychological testing (Volume XX, No. 1), there has been a growing tendency to stress intrinsic test validity and to improve test efficiency. Increasing sophistication with respect to the place of tests for guidance and for diagnostic purposes is evident. The proliferation of projective tests continued despite the paucity of validity data derived from rigorous experimental studies.

The chairman acknowledges the contribution of the chapter authors in the preparation of this issue.

FREDERICK B. DAVIS, *Chairman*

*Committee on Educational and Psychological Testing*

## CHAPTER I

### Testing and the Use of Test Results

FREDERICK B. DAVIS

AMONG the familiar bibliographical sources on tests and their use were the *Third Mental Measurements Yearbook* (6), Swineford and Holzinger's annotated lists of selected references (60), and the February 1950 issue of the *REVIEW OF EDUCATIONAL RESEARCH* (22). Buross is now well along with proofs for the *Fourth Mental Measurements Yearbook*. These year-books are immensely interesting and valuable. The major problem in using them is the variable quality of the reviews.

Goheen and Kavruck (23) assembled a list of 2544 references on how to carry out various aspects of test construction, presumably as a by-product of the preparation of examinations at the U. S. Civil Service Commission. Fuess (21) prepared a history of the College Entrance Examination Board which provides a version of the growth and development of the Board and its testing activities.

#### School Testing Programs

A number of suggestions have been made regarding the nature of testing programs. Diederich (12) discussed the nature of a comprehensive evaluation program, and Bloom (4) described the general plan for the use of examinations in the college of the University of Chicago.

Lindquist (38) urged the use of tests periodically thruout the high-school years to measure all important aspects of each pupil's development. Elicker (14), Erickson (15), Frock (20), and Hastings (31) discussed various aspects of the secondary-school testing program.

Boyer and Eaton (5) wrote on the use of standard tests in Indiana schools, and Segel (51) listed state testing and evaluation programs. Greene and Woodruff (26) linked the improvement of supervision to the use of tests, and Nelson (43) mentioned the fact that community support for the schools can be developed by means of data based on tests and properly presented for laymen.

The use of tests in connection with guidance programs was discussed by many authors. Dressel and Matteson (13) reported an effort to measure three possible effects of the use of test data in counseling. Some evidence suggested that students who participate in interpreting test scores gain more in self-understanding and become more secure in their vocational choices than students who do not so participate. Gustad (29) examined the logic of using test information in counseling and concluded that, properly introduced and used, it is likely to be helpful. Super (59) described two methods of using tests in counseling. In the first, a battery of tests is given at once; in the second, selected tests are used as the need for facts appears in the course of counseling interviews. Super favors the second

method. This problem of the adequate use of tests in counseling was also considered by Percy (45), Rothney (47, 48), Wiener (62), and by Woellner (64).

Failor and Mahler (16) devised a method of checking the adequacy with which tests are selected for use with counselees. Records and tests for use in secondary-school guidance were considered by Roberts and Bauman (46); Harcar and Leonard (30) made specific suggestions for three levels of testing and guidance programs in Catholic secondary schools. Their material is equally relevant for public secondary-school counselors.

Traxler (61) found that from 1941 thru 1951, the median scaled scores of independent secondary-school pupils decreased by .2 to .3 of a standard deviation. The trend was especially noticeable in Spanish and social-studies classes. The median mental ability of the pupils remained the same. Traxler offers some possible reasons for the decline in achievement.

### The Use of Test Scores

Information regarding the use of test scores was published by agencies of three states: California (7), Texas (39), and New York (44). Science Research Associates (50) made available a manual on the use of test results, and four staff members of the Educational Records Bureau prepared an introduction to testing and the use of test results (52). Some practical suggestions for school systems were provided by Cutts (11). Gordon (24) discussed the ways in which tests can be used to secure a better understanding of pupils. Problems in the interpretation of test scores were discussed by Betts (2), Kirk (33), and Schrader (49). Lennon (36) examined the need for improving teachers' understanding of tests.

Bacon (1) explored the reasons for giving tests. Grambs (25) pointed out some ways in which various kinds of situational tests may be used in teacher training, and Wittenborn (63) examined the problem of using the notoriously unreliable difference scores for prediction purposes. Kelly (32) developed a procedure for assigning letter grades (such as A, B, C, D, and E) so that if the variable measured is normally distributed in the population, the mean of each set of letter grades will be an equal distance from its adjacent sets of grades. Bowles (9) made available norms for tests of the College Entrance Examination Board for independent liberal-arts and other types of colleges and for secondary schools.

Kirk (34) deplored the shortcomings of published data about tests and of the representatives (or salesmen) employed by test publishers. Super (58) suggested that test users, plagued by the lack of adequate norms or validation data, develop their own local norms. He thinks that help in accomplishing this might be forthcoming from the test publishers. Stuit (56) discussed at some length the preparation of adequate test manuals.

### Current Evaluation Practices

Michaelis (40) reported the findings of a study during which 100 city school systems were sent questionnaires about their evaluation programs.

Sixty-eight replied, indicating that tests are widely used but that the social and personal characteristics of pupils are not covered by the instruments. Michaelis and Howard (41) analyzed 38 replies to a questionnaire sent to 40 unified school districts with the object of determining how tests and related materials are currently used in school systems.

Findley (17) discussed recent developments in educational evaluation, and Shane (53) reported on such developments with special reference to elementary schools. Ways in which tests are now used were mentioned by Super (57). The relationship of educational objectives and tests was considered by Stanley (54).

### Textbooks

Several textbooks in the field of educational and psychological measurement (excluding statistics texts) have appeared during the last three years. In many respects the most important of these was *Educational Measurement*, edited by Lindquist (37). Sponsored by the American Council on Education and financed by the Grant Foundation, this volume is intended principally for use in graduate courses in educational measurement. The book is divided into three main parts: the Functions of Measurement in Education, the Construction of Achievement Tests, and Measurement Theory. The book and even individual chapters in it have been extensively reviewed and will not be described further in this chapter. It seems to the present writer that thoro acquaintance with the book is necessary for any serious worker in educational and psychological measurement.

At least one of the chapters in the *Handbook of Applied Psychology*, edited by Fryer and Henry, must be mentioned here—the chapter titled "Educational Test Construction" and written by Flanagan (18).

Much of the material in Gulliksen's *Theory of Mental Tests* (27) is sufficiently mathematical in content to be difficult reading except for those who possess considerable mathematical knowledge. Other texts include Cronbach's *Essentials of Psychological Testing* (10), Freeman's *Theory and Practice of Psychological Testing* (19), Stephenson's *Testing School Children* (55), the *Dynamics of Psychological Testing* by Gurvitz (28), and *Measuring Educational Achievement* by Micheels and Karnes (42).

More specialized are Krakower's *Tests and Measurements Applied to Nursing Education* (35), which is a lithoprinted looseleaf book covering basic concepts in measurement with special application to nursing education, and the second edition of Clarke's *Application of Measurement to Health and Physical Education* (8).

### Bibliography

1. BACON, FRANCIS L. "Testing for Testing's Sake." *Journal of the National Education Association* 41: 206-207; April 1952.
2. BETTS, GILBERT L. "Suggestions for a Better Interpretation and Use of Standardized Achievement Tests." *Education* 71: 217-21; December 1950.
3. BLOMMERS, PAUL, chairman. "Methods of Research and Appraisal in Education." *Review of Educational Research* 21: 323-501; December 1951.

4. BLOOM, BENJAMIN S. "Examining: The General Plan." *The Idea and Practice of General Education*. Chicago: University of Chicago Press, 1950. p. 273-81.
5. BOYER, ROSCOE A., and EATON, MERRILL T. *Standardized Testing in the Schools of Indiana*. Bulletin of the School of Education, Vol. 27, No. 1. Bloomington: Indiana University, 1951. 39 p.
6. BUROS, OSCAR K., editor. *Third Mental Measurements Yearbook, 1940-1947*. New Brunswick: Rutgers University Press, 1949. 1047 p.
7. CALIFORNIA STATE DEPARTMENT OF EDUCATION. *Evaluating Pupil Progress*. Bulletin, Vol. 21, No. 6. Sacramento: the Department, 1952. 184 p.
8. CLARKE, H. HARRISON. *Application of Measurement to Health and Physical Education*. Second edition. New York: Prentice-Hall, 1950. 493 p.
9. COLLEGE ENTRANCE EXAMINATION BOARD. *Forty-Ninth Annual Report of the Director*. New York: the Board (425 W. 117th St.) 1949. 83 p.
10. CRONBACH, LEE J. *Essentials of Psychological Testing*. New York: Harper and Brothers, 1949. 475 p.
11. CUTTS, NORMA E. "Use of Tests by the Classroom Teacher." *Measurement and Evaluation in the Improvement of Education*. Report of the Fifteenth Educational Conference, 1950. Washington, D. C.: American Council on Education, 1951. p. 117-20.
12. DIEDERICH, PAUL B. "Design for a Comprehensive Evaluation Program." *School Review* 58: 225-32; April 1950.
13. DRESSEL, PAUL L., and MATTESON, ROSS W. "Effect of Client Participation in Test Interpretation." *Educational and Psychological Measurement* 10: 693-706; Autumn 1950.
14. ELICKER, PAUL E. "Looking at a Testing Program for Secondary-School Youth." *Bulletin of the National Association of Secondary-School Principals* 34: 183-87; May 1950.
15. ERICKSON, ELMER J. "What Kind of Testing Program for Today's Secondary Schools?" *Bulletin of the National Association of Secondary-School Principals* 36: 160-66; April 1952.
16. FAILOR, CLARENCE W., and MAHLER, CLARENCE A. "Examining Counselors' Selection of Tests." *Occupations* 28: 164-67; December 1949.
17. FINDLEY, WARREN G. "Educational Evaluation: Recent Developments." *Social Education* 14: 206-10; May 1950.
18. FLANAGAN, JOHN C. "Educational Test Construction." *Handbook of Applied Psychology*. (Edited by Douglas H. Fryer and Edwin R. Henry) New York: Rinehart and Co., 1950. p. 412-19.
19. FREEMAN, FRANK S. *Theory and Practice of Psychological Testing*. New York: Henry Holt and Co., 1950. 518 p.
20. FROCK, WALTER F. "Basic High-School Testing Program." *Bulletin of the National Association of Secondary-School Principals* 33: 75-80; October 1949.
21. FUESS, CLAUDE M. *The College Board: Its First Fifty Years*. New York: Columbia University Press, 1950. 222 p.
22. GERBERICH, J. RAYMOND, chairman. "Educational and Psychological Testing." *Review of Educational Research* 20: 1-99; February 1950.
23. GOHEEN, HOWARD W., and KAVRUCK, SAMUEL. *Selected References on Test Construction, Mental Test Theory, and Statistics, 1929-1949*. Washington, D. C.: U. S. Civil Service Commission, 1950. 209 p.
24. GORDON, HANS C. "How Teachers Use Test Scores To Understand the Needs of Children." *Growing Points in Educational Research*. 1949 Official Report. Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1949. p. 276-78.
25. GRAMBS, JEAN D. "Some New Examination Patterns in Teacher Education." *Educational Administration and Supervision* 36: 403-10; November 1950.
26. GREENE, JAMES E., and WOODRUFF, WILDA. "A Testing Program for Promoting Improved Supervision." *Educational Administration and Supervision* 35: 346-53; October 1949.
27. GULLIKSEN, HAROLD O. *Theory of Mental Tests*. New York: John Wiley and Sons, 1950. 486 p.
28. GURVITZ, MILTON S. *The Dynamics of Psychological Testing*. New York: Grune and Stratton, 1951. 368 p.
29. GUSTAD, JOHN W. "Test Information and Learning in the Counseling Process." *Educational and Psychological Measurement* 11: 788-95; Winter 1951.

30. HARCAR, GEORGE A., and LEONARD, REGIS J. "Minimum Guidance Testing Programs for Catholic Secondary Schools." *Catholic Educational Review* 50: 394-402; June 1952.
31. HASTINGS, J. THOMAS. "What Kind of Testing Program in Today's Secondary School?" *Bulletin of the National Association of Secondary-School Principals* 36: 154-60; April 1952.
32. KELLEY, TRUMAN L. "The Use of Literal Grades." *Journal of Educational Psychology* 41: 488-92; December 1950.
33. KIRK, BARBARA A. "Individualizing of Test Interpretation." *Occupations* 30: 500-505; April 1952.
34. KIRK, BARBARA A. "Test Distributors and Our Needs." *Occupations* 29: 257-59; January 1951.
35. KRAKOWER, HYMAN. *Tests and Measurements Applied to Nursing Education*. New York: G. P. Putnam's Sons, 1949. 179 p.
36. LENNON, ROGER, T. "Needed Improvement of Teachers' Understanding of Tests." *Growing Points in Educational Research*. 1949 Official Report. Washington, D. C.: American Educational Research Association, a department of the National Education Association, 1949. p. 271-75.
37. LINDQUIST, EVERET F., editor. *Educational Measurement*. Washington, D. C.: American Council on Education, 1951. 819 p.
38. LINDQUIST, EVERET F. "Some Criteria of an Effective High-School Testing Program." *Measurement and Evaluation in the Improvement of Education*. Report of the Fifteenth Educational Conference, 1950. Washington, D. C.: American Council on Education, 1951. p. 17-33.
39. MANUEL, HERSCHEL T. *Testing and Test Results*. Austin: Texas Commission on Coordination in Education, 1949. 27 p.
40. MICHAELIS, JOHN U. "Current Practices in Evaluation in City School Systems." *Educational and Psychological Measurement* 9: 15-22; Spring 1949.
41. MICHAELIS, JOHN U., and HOWARD, CHARLES. "Current Practices in Evaluation in City School Systems in California." *Journal of Educational Research* 43: 250-60; December 1949.
42. MICHEELS, WILLIAM J., and KARNES, M. RAY. *Measuring Educational Achievement*. New York: McGraw-Hill Book Co., 1950. 496 p.
43. NELSON, LESTER W. "Use of Tests in School Administration." *Measurement and Evaluation in the Improvement of Education*. Report of the Fifteenth Educational Conference, 1950. Washington, D. C.: American Council on Education, 1951. p. 113-16.
44. NEW YORK STATE EDUCATION DEPARTMENT. *School Testing Program: A Guide to the Selection and Use of Standardized Tests*. Bulletin No. 1397. Albany: the Department (University of the State of New York), 1950. 24 p.
45. PERCY, MILDRED S. "Use of Tests in the Guidance Program of Public Schools." *Measurement and Evaluation in the Improvement of Education*. Report of the Fifteenth Educational Conference, 1950. Washington, D. C.: American Council on Education, 1951. p. 121-23.
46. ROBERTS, JOHN R., and BAUMAN, MARY K. "Records and Tests for Guidance in the Secondary School." *Education* 70: 311-21; January 1950.
47. ROTHNEY, JOHN W. M. "Interpreting Test Scores to Counselees." *Occupations* 30: 320-22; February 1952.
48. ROTHNEY, JOHN W. M. "Techniques in Studying Individuals." *High School Journal* 33: 216-19; December 1950.
49. SCHRADER, WILLIAM B. "Making Test Scores Meaningful." *College Board Review* 15: 202-208; May 1951.
50. SCIENCE RESEARCH ASSOCIATES. *How To Use the Test Results*. Chicago: Science Research Associates, 1949. 32 p.
51. SEGEL, DAVID. *State Testing and Evaluation Programs*. Circular No. 320. Washington, D. C.: U. S. Office of Education, Federal Security Agency, 1951. 38 p.
52. SELOVER, MARGARET, and OTHERS. *Introduction to Testing and the Use of Test Results*. Educational Records Bulletin No. 55. New York: Educational Records Bureau, 1950, 107 p.
53. SHANE, HAROLD G. "Recent Developments in Elementary-School Evaluation." *Journal of Educational Research* 44: 491-506; March 1951.
54. STANLEY, JULIAN C. JR. "Standardized Tests and Educational Objectives." *Peabody Journal of Education* 28: 218-21; January 1951.

55. STEPHENSON, WILLIAM. *Testing School Children*. New York: Longmans, Green and Co., 1949. 127 p.
56. STUIT, DEWEY B. "Preparation of a Test Manual." *American Psychologist* 6: 167-70; May 1951.
57. SUPER, DONALD E. "Current Trends in the Use of Tests." *American Vocational Journal* 25: 13-14; April 1950.
58. SUPER, DONALD E. "Dilemma for Test Users." *Occupations* 29: 174-76; December 1950.
59. SUPER, DONALD E. "Testing and Using Test Results in Counseling." *Occupations* 29: 95-97; November 1950.
60. SWINEFORD, FRANCES, and HOLZINGER, KARL J., compilers. "Selected References on Statistics, the Theory of Test Construction, and Factor Analysis." *School Review* 58: 489-93; November 1950. 59: 489-97; November 1951.
61. TRAXLER, ARTHUR E. "Trends in Achievement of Independent Secondary-School Pupils During a Ten-Year Period." *1951 Achievement Testing Program in Independent Schools, and Supplementary Studies*. Educational Records Bureau Bulletin No. 57. New York: Educational Records Bureau, 1951. p. 67-78.
62. WIENER, FREDERICK. "Use of Tests in Guidance." *Occupations* 30: 662-63; May 1952.
63. WITTENBORN, JOHN R. "Evaluation of the Use of Difference Scores in Prediction." *Journal of Clinical Psychology* 7: 108-11; April 1951.
64. WOELLNER, ROBERT C. "Interpretation of Test Results in Counseling." *School Review* 59: 515-17; December 1951.

## CHAPTER II

### Development and Applications of Tests of General Mental Ability

JULIAN C. STANLEY\*

BECAUSE "intelligence" tests permeate most areas of education and psychology, the writer has found it imperative arbitrarily to omit much clinical material from this chapter and to treat the literature on individual differences very lightly in order to prevent his bibliography from preempting all space allotted for comments. Thus a number of important and interesting studies in related areas get little or no recognition here.

#### General Overview

During the three years since Cornell and Gillette (64) reported on 63 studies there seem to have been few basic changes but myriad extensions and refinements. For example, the long-held practice of employing a group intelligence test distinct from measures of aptitude appears threatened by the *Differential Aptitude Test* (DAT). Williams (285) obtained an  $r$  of .73 between the DAT Verbal Reasoning subtest and IQ's on Form L of the *Revised Stanford-Binet Intelligence Scale* for 50 high-school sophomore white girls, .55 for the DAT Abstract Reasoning subtest with the S-B, and .78 for the DAT Verbal and *Henmon-Nelson*. While not high enough to denote interchangeableness, these figures do indicate considerable common variance.

Correlations between the eight DAT subtests and seven group tests of intelligence were in general so substantial that Bennett, Seashore, and Wesman (27) deemed it unnecessary to employ an intelligence test of the usual type when DAT results are available.

The studies of Millard (186) and Wickert (284) showed that *How Supervise?* functions somewhat as an intelligence test for persons who did not complete high school but as a measure of supervisory knowledge for relatively well-educated individuals; Levine's article (169) is pertinent here. Gurvitz (124) found the *Revised Minnesota Paper Form Board Test* related more to intelligence and general cultural level than to mechanical ability, its correlation with *Army Alpha* scores being .685. Thus there is no clear dichotomy of intelligence tests versus aptitude or achievement tests.

#### Books

Goodenough (114) provided an excellent history of mental testing which also contains much methodological and theoretical material. Kent (153)

\* Assisted with bibliographic and secretarial details by Margaret T. Aldridge and Doris Roberts.

emphasized qualitative aspects of mental testing, as did Stephenson (246). Vernon (270) published a concise, integrated summary of factor analysis studies; two of his chapters are devoted directly to intelligence. Cronbach (68) and Freeman (101) gave considerable attention to individual intelligence tests in their elementary textbooks. Super (249) concluded that for vocational guidance purposes group intelligence tests were at least as useful as the *Wechsler-Bellevue Intelligence Scales* and certainly more economical.

### Theoretical Articles

Jastak (143, 144, 145) and Cassel (53) discussed criteria of feeble-mindedness, the former proposing an "altitude quotient" based upon the individual's highest ability. Apparently Jastak's heuristic analyses call for greater reliability than is likely to be found in most clinical testing. Errors of measurement, covered thoroly by Gulliksen (119), may jeopardize Jastak's "rigorous criterion" (144). Various approaches to theories of intelligence were explored by Arthur (19), Combs (62), Hick (135), Knehr (155), Raven (214), and Wechsler (278, 280).

### Longitudinal Studies

In a 32-page article, Bayley (26) discussed factors of variability and consistency for 41 children tested repeatedly from one month thru 18 years of age. She found high  $r$ 's between scores on the *Stanford-Binet*, *Wechsler-Bellevue*, and *Terman-McNemar* tests. Knehr and Sobol (156) did not discover significant IQ differences between 99 prematurely born children and a control group during the early school years. Writing in the methodologically controversial and complicated area of foster-home influences on intelligence, Skodak and Skeels (232) offered a comprehensive analysis of their long-range study, and concluded that the adopted children performed consistently better on intelligence tests than would have been predicted from available data concerning their true parents, and that they equaled or surpassed "the mental level of own children in environments similar to those which have been provided by the foster parents." Skodak (231) dealt with IQ resemblances of unrelated adopted children in the same family. Richards (215) felt that fluctuations in the IQ of the one child whom he studied were closely related to the current life situation.

Swanson (250) found that intelligence-test gains after 20 years were much greater for a college graduate group than for nongraduates and non-attenders. Pressey (211) summarized numerous studies showing that the *Ohio State University Psychological Test* was a valuable aid in deciding which college freshmen should be accelerated.

Escalona (91) supplemented her theoretical article concerning the predictive value of infant tests with empirical evidence suggesting that those infants who in the opinion of the examiner at the time of the initial test functioned optimally show less discrepancy on a retest than those who functioned less well.

### The 1947 Repetition of the 1932 Scottish Survey

The failure of Scottish 11-year-olds to decline in verbal intelligence cross-sectionally from 1932 to 1947 as predicted on the basis of the definitely negative correlation between the size of families and the intelligence of children therein (257) aroused stimulating discussions by Burt (46, 47), Penrose (202, 203), Thomson (255, 256), and Vernon (268, 269). Cattell (55) retested 10-year-olds in England with a nonverbal intelligence test after a lapse of 13 years, also finding an over-all slight but significant increase in IQ. As possible causes, the various writers mentioned practice effect, inadequacies of the tests used, differential migration, heightened environmental stimulation, and a self-stabilizing genetical system. Articles on related topics were 20, 149, 229. The Scottish Survey material seems to have highly important implications for intelligence-test theory and practice.

### Factor Analyses and Other Correlational Studies

General factors continued to be studied. Rimoldi (216) identified his second-order unrotated general factor as Spearman's  $g$ . Ingham (142) considered that a factor other than  $g$  was needed to explain the intercorrelations among eight memory tests. Curtis (70), Doppelt (81), Hagen (125), and Swineford (251, 252) found no tendency for the general factor to decrease in importance among children with age. They were essentially in agreement with the trend noted three years ago by Cornell and Gillette (64).

Allen and Bessell (4) reported that the *Alpha Form 9*, *Otis Quick-Scoring*, and *Henmon-Nelson* tests intercorrelated an average of .71, compared with their average  $r$  of .34 with the *Chicago Non-Verbal Examination*. Bailey's investigation (21) of the intercorrelations and predictive value of several intelligence tests led to a local adoption of the *California Short-Form Test of Mental Maturity*, Primary and Elementary Forms.

The well-organized study by Heil and Horn (132) revealed considerable norm and validity differences among the *Otis Self-Administering* (Form A), *California Short-Form*, *SRA Primary Mental Abilities*, *SRA Non-Verbal*, and *Terman-McNemar* tests. Correlations with five-semester grade-point averages were: PMA total, .46; *Terman-McNemar*, .46; CTMM, .41; *Otis*, .39; and SRA, .32. The mean PMA IQ's were quite low compared with the other tests, the SRA and CTMM mean IQ's quite high. In general, the *Terman-McNemar* was judged most satisfactory.

Garrett (106) dealt comprehensively with factors related to college success, analyzing 194 studies and concluding that high-school scholarship is the best predictor (.56), with general achievement tests and intelligence tests next (.49 and .47). Rosilda (219), using data secured under heterogeneous testing conditions, obtained an  $r$  of only .42 between CTMM IQ's and percentile ranks on a standardized algebra achievement test,  $N$  being 635. Lehman (161) found no significant correlation between *Otis* IQ's and gains on a music test. Tho, in their first study, Lorge and Kruglov (173) did not find the readability of compositions significantly related to

intelligence, later (172) they obtained significant  $r$ 's of .47 for readability and .70 for rated merit.

Having examined critically those deficiencies in the learning criterion which frequently have resulted in low correlations between achievement gains and intelligence, Tilton (261) used improved measures to secure two  $r$ 's of .49. A factor analysis by Tilton, scheduled for publication in the January 1953 *Journal of Psychology*, is fairly consistent with the view that there is a general ability to learn which can be identified with the "general" intelligence test. Smith (236) also decided that there is a positive correlation between learning gain and intelligence.

Davenport and Remmers (73) carried out a factor analysis of correlations between state means on the *A-12 V-12 Examination*, administered to 300,000 servicemen in 1943, and 13 state characteristics, finding "state economic," "rural-urban," and "deep-South versus non-South" factors. The four most valid variables yielded a multiple  $r$  of .962. For 154 communities Thorndike (258) found the partial  $r$  between *Pintner Intelligence Tests (Verbal Series)* IQ's and *Metropolitan Achievement Test* scores in Grades II-IX, with age held constant, to be .67. Using 24 community variables from 1940 census data he estimated maximum multiple  $r$ 's to be approximately .55 to .60 for intelligence but only .30 for achievement. Several possible hypotheses concerning this discrepancy were presented.

### Physical and Environmental Factors

Special considerations in the testing of cerebral-palsied children were discussed by Holden (137), Jewell and Wursten (146), and Tracht (263). Berlinsky (31) reviewed the literature concerning the intelligence of the deaf and concluded that this group averages slightly lower than nondeaf individuals, the age of onset seeming to make no difference. Hayes (129, 130) contributed two chapters on measuring the intelligence of the blind. Sloan (234) found motor proficiency positively related to intelligence.

The long-awaited report by Eells and others (85) concerning socioeconomic influences on intelligence-test performance appeared in 1951. Many professional persons will find this volume interesting but will want to read the not-particularly-favorable reviews by Darley (72) and McNemar (176).

Gellerman and Hays (108) attempted to devise a measure of cultural knowledge uncorrelated with intelligence and concluded that this is possible. About one-third of Educational Testing Service's 1949 conference (45) was devoted to the "Influence of Cultural Background on Test Performance," with papers by Anastasi, Haggard, Stephenson, and Turnbull. Gurvitz (120) found that much of the apparent decline in intelligence of male prisoners with age was due to unequal educational opportunities and occurred at the low IQ levels. The smaller mean postwar IQ of boys entering a Dutch industrial school was attributed by de Groot (78) to disrupting effects of World War II upon the extent and quality of education.

In a study which attempted to control relevant variables, Carlson and

Henderson (50) confirmed the usually reported substantial superiority of white non-Mexican children over Mexican ones on verbal intelligence tests, but found a similar nonverbal discrepancy on the *California Test of Mental Maturity*. This is in conflict with Darcy's difference (71), using Pintner tests, of eight points in favor of the mean nonverbal IQ for 235 children of Puerto Rican parentage.

The continued facilitating effect of repeated testing and its positive correlation with intelligence level were established by Cane and Heim (49) in four experiments. Retest practice effects were also found by Peel (201) and Rudolf (220). Berk (30) discovered a considerable amount of intelligence-test coaching in an institution for mentally defective delinquents.

### Specific Tests and Their Applications

In response to letters sent to the major test companies, the writer was deluged with valuable information, much of it as yet unpublished. Unfortunately, he is able to use little of it here because of space limitations. Generally, in the condensed summary that follows, only newly published tests and really major revisions of old ones are mentioned.

Almost surely the most enthusiastically received new test during the three-year period was the *Wechsler Intelligence Scale for Children* (281), abbreviated WISC, a downward extension and restandardization of the *Wechsler-Bellevue Intelligence Scale, Form II*. Recommended for ages 5 thru 15, it thus overlaps with the W-B I and II at ages 10 thru 15, and like them yields Verbal, Performance, and Full-Scale deviation IQ's. Various aspects of extensive standardization data have been reported by the following: Hagen (125); Krugman and others (158); Seashore (224); Seashore, Wesman, and Doppelt (225); and Wechsler (281). There is some evidence that the mean WISC P IQ is higher than the V IQ (59, 79, 98, 117, 199, 234, 237), tho contradictory studies are not lacking (158, 222, 282, 289). The S-B IQ has been found in several instances (59, 98, 158, 199, 282) to exceed the WISC FS IQ, except for the mentally deficient (189, 222, 234, 237). Other published WISC articles (5, 118, 279) make the total number to date 19.

The *Leiter-Partington Adult Performance Scale* (67, 163, 164, 166, 195, 196, 274), a painted-cube test which is an adaptation of both Arthur's *Stencil Design Test* and the *Partington Pathways Test*, was designed to be a measure of general intelligence also useful for clinical and diagnostic purposes. It is independent of the carefully constructed *Leiter International Performance Scale* (32, 34, 162, 165, 183, 254), abbreviated LIPS, which dates back to 1940. The *Arthur Adaptation of the Leiter International Performance Scale* (17) is to be used along with Arthur's *Point Scale of Performance Tests, Revised Form II* for children of CA 4.00 to 7.99 or having MA's within that range; the AALIPS goes down to 3.00. Wholly untimed, it is given without verbal instructions and should be useful for testing young children with physical and linguistic handicaps.

Gilliland's *Northwestern Intelligence Tests, Forms A* (4 to 12 weeks)

and B (13 to 36 weeks) (110, 111, 112) each consist of 40 developmental-response items and yield IQ's.

Several promising new group measures appeared. The *Kuhlmann-Finch Intelligence Tests* (92) were offered as an adequately prepared sequel to the *Kuhlmann-Binet* individual test; the *Kuhlmann-Anderson Intelligence Test* (159) in its sixth edition remains on the market. The K-F tests consist of eight separate nonoverlapping booklets, each containing five subtests, for Grades I, II, III, IV, V, VI, junior high, and senior high. Cultural influences have been minimized and sex differences virtually eliminated. Reliability data are especially complete.

For many years Holzinger has been conducting factor analyses and contributing to the theory of intelligence. Now on the market are the *Holzinger-Crowder Uni-Factor Tests* (139), two comparable forms for Grades VII thru XII that contain verbal, spatial, numerical, and reasoning subtests.

The pictorial *Davis-Eells Games* (75, 76), designed for Grades I and II, and III thru VI, are meant to be culture-fair. Manuel's *Cooperative Inter-American Tests* (179, 180, 181) included 12 general-ability tests, culture-equated comparable forms for primary, intermediate, and advanced levels that were constructed simultaneously in English and Spanish.

Goossen's (115) ingeniously disguised six-item intelligence test proved quite valid and feasible for public-opinion surveys where an estimate of the mental level of each respondent was desired. Hanna's (127) interview estimates of intelligence correlated .71 with ACE *Psychological Examination* scores and .66 with the *Ohio State University Psychological Test*. The ACEPE and OSUPT correlated .77. Engle and Hamlett (89, 90) considered the 10-minute *Buck Time Appreciation Test* highly enough correlated with the *Revised S-B* (.65) to serve as a screening or supplementary test for mentally deficient patients and sufficiently reliable over a three-year test-retest interval (.82), tho it tended to yield higher MA's and IQ's than the S-B. Semeonoff and Laird (226) were only partially successful in obtaining a valid intelligence score from the *Vigotsky Test*.

An item-analyzed short form of the *Otis Alpha* is now available (193). The *Thurstone Test of Mental Alertness* has been revised completely and published in three comparable forms (259).

Since the introduction of the *Full-Range Picture Vocabulary Test* in 1949, Ammons and his collaborators (9, 10, 11, 12, 13, 14, 63) have reported on six different norm groups and concluded that it is essentially an intelligence test.

### The Wechsler-Bellevue Intelligence Scales, Forms I and II

Rabin and Guertin (212) reviewed W-B research from 1945 until about June 30, 1950. Their 145-item bibliography contains 28 references that appeared in 1949 and 26 for 1950. These will not be duplicated here.

Burton (48) found that in psychological clinics the two most frequently used intelligence tests were the W-B and the S-B, in that order. Gurvitz

(122) criticized several aspects of the W-B I manual rather severely, with particular attention to Tables 39, 40, and 41. Block, Levine, and McNemar (36) outlined a modified triple-classification analysis-of-variance design useful for detecting the existence of psychometric patterns which differentiate various clinical groups by testing the group  $\times$  variable interaction for significance. Kitzing and Blumberg (154) provided brief supplementary instructions for administering the W-B I and for scoring the more troublesome responses.

Gerboth (109) compared W-B I and II results for superior college students, Hays and Schneider (131) for mental defectives. They found overall similarity but subtest discrepancies. Steisel (244, 245) reported significant retest gains. Webb and De Haan (275, 276) and Helmick (133) argued about split-half reliabilities and variability among normals versus schizophrenics.

Bensberg and Sloan (28) cast doubt on Wechsler's standardization sampling of older mental defectives and his concept of "normal deterioration" at this intelligence level. Fox and Birren (96) found normal whites 60 to 69 years of age highest on Information, Vocabulary, and Comprehension and lowest on Block Design, Picture Arrangement, and Digit Symbol, in close agreement with the results of other investigations. Gurvitz (123) attributed performance decrement with age to loss of speed rather than quality. The studies of Cohen (60, 61), Davis (77), and Wittenborn and Holzberg (286) directly or by implication constitute a serious challenge to the mechanical use of the W-B as an aid in clinical diagnosis.

Scherer (223) discovered that 22 mental patients performed significantly better on the Digit Symbol test in an individual testing situation than in a group setting. Davidson and others (74) found whites higher on  $P$  than  $V$  but Negroes lower on  $P$ . Webb and Haner (277) demonstrated the possibility of scoring the W-B I Vocabulary subtest more quantitatively. Stacey and Portnoy (239), and Stacey and Markin (238) concluded that the descriptive method of concept formation seems to be a higher or more complex level than the functional method. Various methodological problems were attacked by Alimena (3), Burik (43), Eglash (86), Newton (190), and Shannon and Rossi (227).

Alderdice and Butler (2) obtained an  $r$  of .80 between W-B I  $V$  and S-B L IQ's for a mentally defective group whose SD on either scale the writer estimates to be only 9. Frandsen (97) found both the W-B  $FS$  and  $V$  IQ's better correlated with high-school grades (.69) than was the *Hennon-Nelson* (.52). Storrs (248) secured an  $r$  of .80 between W-B  $V$  IQ's and the  $G$  test of the *USES General Aptitude Test Battery*.

Various short forms of the W-B will be mentioned later in this review.

### The Revised (1937) Stanford-Binet Intelligence Scales

Jones' (148) orthogonal centroid factor analysis of Terman-Merrill standardization data for age levels 7, 9, 11, and 13 revealed varying group factors at the four levels but no general factor. Aborn and Derner (1),

Baldwin (23), and Roberts and Mellone (217) showed that the markedly different standard deviations reported by Terman and Merrill for several age levels are attributable to unequal item difficulties at these age levels rather than to accidents of standardization sampling. Roberts and Mellone described refined procedures for correcting IQ's within the age range 5-0 to 14-11 and also discussed the possible influence of differential skewness. Elwood (88) found slight mean IQ changes in three retarded preprimary groups over a two-year period.

On the basis of research findings Frandsen, McCullough, and Stone (99) endorsed serial-order administration of S-B tests and interpretation of resulting IQ's in the usual manner. Pierce (205) gave appropriate advice concerning common errors in S-B administration. Gordon and Durea (116) and Sacks (221) produced experimental evidence concerning, respectively, the deleterious influence of discouragement upon retests and the effects of child-examiner contacts outside the testing situation.

Baldwin (22) and Magaret and Thompson (177) showed that bright children answered correctly more "intellectual" items than normal or dull children. Bond and Fay (38) obtained similar results with good versus poor readers matched for MA.

Cruikshank and Qualtere (69) found an  $r$  of .90 between scores on the original (1916) S-B and the *Revised S-B, Form L*. Tho the respective IQ means were 71.98 and 70.19, the difference between them was highly significant.

For 27 imbeciles Pascal and others (197) reported a rho of .61 between S-B MA and ability to delay an instrumental response leading to reward. Elonen (87) compared S-B and *Kuhlmann Tests of Mental Ability* scores for six varied groups and found the S-B mean greater for all except the high-IQ student group.

### Other Intelligence Tests for Children

Arthur (18) found approximately the same median IQ's for 60 "simple aments" tested with her *Point Scale of Performance, Form I* and the S-B. Gellerman (107) suggested a restandardization of Arthur's Form II, citing wide differences between I and II. Hamilton (126), Johnson (147), and Manolakes and Sheldon (178) disclosed large discrepancies between the S-B and Form II norms.

Birch (33) recommended the Goodenough *Draw-a-Man Test* as a valid measure of mental ability for children of S-B IQ 70 or lower with CA's between 10-6 and 16-3, in addition to its customary use with younger children. Stonesifer (247) was not able by use of the test to differentiate schizophrenic from nonpsychotic subjects matched for age and education.

Ansbacher (16) found the *Draw-a-Man Test* less closely correlated with Thurstone's *Primary Mental Abilities Test* (PMA) Verbal Meaning score (.26) than with Reasoning (.40), Space (.38), and Perception (.37). Smith (235) obtained an  $r$  of .78 between W-B and PMA IQ's, but the PMA mean was 7.2 points lower than the S-B mean. Ramaseshan (213) matched bright

and dull ninth-graders for PMA MA and found the bright group significantly better on Verbal Meaning and Reasoning but significantly inferior on Space and Word Fluency. McKee (174) deemed the PMA adequate for testing superior five-year-olds and all but very superior six-year-olds, though in most cases it yielded slightly lower scores than the S-B.

### "Culture-Free" Tests

Tilton (260) discovered that scores on the *Cattell Culture-Free Test* correlated .84 with W-B IQ's much higher than with either the *Otis Group Examination* or the *Henmon-Nelson*. Pierce-Jones and Tyler (206) found it a poorer predictor of scores on two psychology examinations than were *Q*, *L*, or *T* scores of the *ACE Psychological Examination*. Cattell (54) cited evidence that as a test becomes freer from scholastic contamination the standard deviation of IQ's virtually doubles.

Cassell (52), Foulds and Raven (95), Keir (150), Notcutt (191), and Sinha (230) published studies dealing with *Raven's Progressive Matrices Test*. Porteus (208, 209, 210) and Tizard (262) reported on research with the *Porteus Maze Test*.

### Other Tests

As usual, the *ACE Psychological Examination for College Freshmen* was employed widely in prediction studies (25, 29, 42, 105, 192, 266, 272), especially with regard to the differential predictive value of its *Q* and *L* scores (37, 39, 51, 58, 100, 265, 273). Other reports concerned its correlation with tests of critical thinking (104), improvement in scores during college (228), and equating five forms of the high-school version (15).

Investigations involving the *Army General Classification Test* were conducted by Altus (6), Fulk and Harrell (103), and Tamminen (253). Four reports (40, 41, 128, 204) dealt with the *Armed Forces Qualification Test* (AFQT). Pastore (198) commented on the inadequacy of the *Army Alpha* and *Beta* tests as bases for comparing the intelligence of whites and Negroes.

More than 339,000 persons took the *Selective Service College Qualification Test* (SSCQT) during the spring and summer of 1951. The background of this test was set forth by Findley (93). Two comprehensive reports of sectional and academic area differences (56, 84) placed the East-South-Central region and education majors lowest, with the Middle Atlantic region and engineering students highest.

Problems related to the supply, identification, and conservation of high-level intellectual talent were explored by Wolfe (287), Wolfe and Oxtoby (288), and Dyer's symposium (82).

The restricted *Miller Analogies Test* (187), three forms of which are available for scholastic prediction among graduate students, was studied by Blake (35), Doppelt (80), Glaser (113), Kelly and Fiske (151), Stafford (240), and Zagorski (290). Levine's *Minnesota Psycho-Analogies Test* (168, 170, 200) seems to be a promising instrument for use in the selection of graduate psychology students and MA-level psychol-

ogists. Travers and Wallace (264) devised an *Academic Aptitude Test, Graduate Level* which appeared to be more valid than the *Miller Analogies Test* in four out of five subject areas. Wallace (271) reported that lecturers and research workers did better than advanced students on Heim's *AH5 Test*. Lannholm and Schrader (160) found "satisfactory" validities for the Verbal Factor Profile Test of the *Graduate Record Examinations* in English, history, and social studies but lower  $r$ 's in other fields. Roe (218) administered a specially devised verbal-spatial-mathematical test to 61 eminent scientists.

Altus and Altus (7) and Altus and Thompson (8) found the incidence of unstereotyped human movement responses on Monroe's *Group Rorschach* highly reliable and substantially correlated curvilinearly with intelligence. Burnham (44) and Holzberg and Belmont (138) reported low, insignificant correlations between various *Rorschach* and W-B factors.

### Brief Measures of Intelligence

The perennially popular quest for shorter tests continued. Mensh (185) provided a comprehensive review of the rationale for these. McNemar (175), Herring (134), and Hilden and Taylor (136) compared various short forms of the W-B, and Knott and others (157) discovered substantial relationships between several of the abbreviated Kent tests and the W-B. Other brief W-B's were offered by Cotzin and Gallagher (66), Finkelstein, Gerboth, and Westerhold (94), and Gurvitz (121). Meister and Kurko (184) dealt with a shortened S-B.

Corsini's *Immediate Test* (65), a vocabulary-age scale requiring only 31½ minutes for administration and scoring and consisting chiefly of concrete nouns, correlated .77 to .90 with the *Otis*, S-B, and W-B. Otis and Chesler (194) introduced the 10- or 15-minute *Classification Test for Industrial and Office Personnel*, Forms A and B, containing 100 verbal items of approximately uniform difficulty. Chesler (57) and Lindzey (171) discussed the *Wonderlic Personnel Test*.

Hunt and French (140) developed the *Navy-Northwestern Matrices Test* (NNMT), a brief nonverbal measure designed to correlate well with standard verbal tests and to be useful diagnostically. The *CVS Abbreviated Individual Intelligence Scale* (102, 141, 182, 207) on which they have worked for several years consists of the W-B Comprehension and Similarities subtests, together with a 15-word vocabulary test which Thorndike adapted from the S-B.

### Miscellaneous

Dyer (83) reported on continuing research with the *Scholastic Aptitude Test* (SAT) of the College Entrance Examination Board, which is taken by approximately 70,000 candidates each year. Traxler (267) summarized experience derived from the administration of the *Junior Scholastic Aptitude Test* (JSAT) of the Educational Records Bureau to 60,243 private-school students.

Mursell (188) described a simplified case for the *Kuhlmann Scale of Mental Development*. Lennon (167) provided equivalent scores and IQ's for certain *Otis Quick-Scoring*, *Pintner Verbal*, and *Terman-McNemar* forms.

Steele's questionnaire (243) revealed that the intelligence tests most frequently used by employers in the selection of college graduates were the *Wonderlic* and the *Otis*. Kenney (152) found that 20 percent of the items in high-school level intelligence tests are mathematical and that many of these could have been taken directly from mathematics textbooks. Barbe and Grilk (24), Stanley (242), and Wheeler (283) published  $r$ 's of .72, .80, and .71, respectively, between reading and intelligence-test total scores for quite diverse groups.

### Concluding Remarks

There is considerable need for more careful planning of investigations, greater sophistication in test theory (especially with regard to errors of measurement), and better grasp of statistical procedures, including the analysis of variance and covariance. Since correlational technics are fundamental to the entire area, each psychometric researcher should have a *thoro* knowledge of such matters as attenuation and restriction of range. This can hardly be acquired in the usual elementary measurement or statistics course, so advanced training seems imperative (241).

### Bibliography

1. ABORN, MURRAY, and DERNER, GORDON F. "IQ Variability in Relation to Age on the Revised Stanford-Binet." *Journal of Consulting Psychology* 15: 231-35; June 1951.
2. ALDERDICE, ERNEST T., and BUTLER, A. J. "An Analysis of the Performance of Mental Defectives on the Revised Stanford-Binet and the Wechsler-Bellevue Intelligence Scale." *American Journal of Mental Deficiency* 56: 609-14; January 1952.
3. ALIMENA, BENJAMIN S. "Norms for Scatter Analysis on the Wechsler Intelligence Scale." *Journal of Clinical Psychology* 7: 289-90; July 1951.
4. ALLEN, ROBERT M., and BESSELL, HAROLD, "Intercorrelations among Group Verbal and Non-Verbal Tests of Intelligence." *Journal of Educational Research* 43: 394-95; January 1950.
5. ALTUS, GRACE T. "A Note on the Validity of the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 16: 231; June 1952.
6. ALTUS, WILLIAM D. "The Height and Weight of Soldiers in Association with Scores Earned on the Army General Classification Test." *Journal of Social Psychology* 29: 201-10; May 1949.
7. ALTUS, WILLIAM D., and ALTUS, GRACE T. "Rorschach Movement Variables and Verbal Intelligence." *Journal of Abnormal and Social Psychology* 47: 531-33; April 1952. (Supplement)
8. ALTUS, WILLIAM D., and THOMPSON, GRACE M. "The Rorschach as a Measure of Intelligence." *Journal of Consulting Psychology* 13: 341-47; October 1949.
9. AMMONS, ROBERT B., and ACUERO, ABELARDO. "The Full-Range Picture Vocabulary Test: VII. Results for a Spanish-American School-Age Population." *Journal of Social Psychology* 32: 3-10; August 1950.
10. AMMONS, ROBERT B.; ARNOLD, PAUL R.; and HERRMANN, ROBERT S. "The Full-Range Picture Vocabulary Test: IV. Results for a White School Population." *Journal of Clinical Psychology* 6: 164-69; April 1950.

11. AMMONS, ROBERT B., and HOLMES, JAMES C. "The Full-Range Picture Vocabulary Test: III. Results for a Preschool-Age Population." *Child Development* 20: 5-14; March 1949.
12. AMMONS, ROBERT B.; LARSON, WILLIAM L.; and SHEARN, CHARLES R. "The Full-Range Picture Vocabulary Test: V. Results for an Adult Population." *Journal of Consulting Psychology* 14: 150-55; April 1950.
13. AMMONS, ROBERT B., and MANAHAN, NEIL. "The Full-Range Picture Vocabulary Test: VI. Results for a Rural Population." *Journal of Educational Research* 44: 14-21; September 1950.
14. AMMONS, ROBERT B., and RACHIELE, LEO D. "The Full-Range Picture Vocabulary Test: II. Selection of Items for Final Scales." *Educational and Psychological Measurement* 10: 307-19; Summer 1950.
15. ANGOFF, WILLIAM H. "Equating of the ACE Psychological Examinations for High School Students." (Abstract) *American Psychologist* 7: 287; July 1952.
16. ANSBACHER, HEINZ L. "The Goodenough Draw-a-Man Test and Primary Mental Abilities." *Journal of Consulting Psychology* 16: 176-80; June 1952.
17. ARTHUR, GRACE. *The Arthur Adaptation of the Leiter International Performance Scale*. Washington, D. C.: Psychological Service Center Press, 1952. 73 p.
18. ARTHUR, GRACE. "The Relative Difficulty of Various Tests for Sixty Feeble-minded Individuals." *Journal of Clinical Psychology* 6: 276-79; July 1950.
19. ARTHUR, GRACE. "Some Factors Contributing to Errors in the Diagnosis of Feeble-mindedness." *American Journal of Mental Deficiency* 54: 495-501; April 1950.
20. ASPELL, S. A. "The Relation Between Inbreeding and Intelligence." *Human Biology* 20: 171-81; December 1948.
21. BAILEY, HELEN K. "A Study of the Correlations Between Group Mental Tests, the Stanford-Binet, and the Progressive Achievement Test Used in the Colorado Springs Elementary Schools." *Journal of Educational Research* 43: 93-100; October 1949.
22. BALDWIN, ALFRED L. "The Relative Difficulty of Stanford-Binet Items and Their Relation to I.Q." *Journal of Personality* 16: 417-30; June 1948.
23. BALDWIN, ALFRED L. "Variation in Stanford-Binet I.Q. Resulting from an Artifact of the Test." *Journal of Personality* 17: 186-98; December 1948.
24. BARBE, WALTER, and GRILK, WERNER. "Correlations Between Reading Factors and I.Q." *School and Society* 75: 134-36; March 1, 1952.
25. BARRETT, DOROTHY M. "Differential Value of Q and L Scores on the ACE Psychological Examination for Predicting Achievement in College Mathematics." *Journal of Psychology* 33: 205-207; April 1952.
26. BAYLEY, NANCY. "Consistency and Variability in the Growth of Intelligence from Birth to Eighteen Years." *Journal of Genetic Psychology* 75: 165-96; December 1949.
27. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *A Manual for the Differential Aptitude Tests*. Second edition. New York: Psychological Corp., 1952. 77 p.
28. BENSBERG, GERARD J., and SLOAN, WILLIAM. "A Study of Wechsler's Concept of 'Normal Deterioration' in Older Mental Defectives." *Journal of Clinical Psychology* 6: 359-62; October 1950.
29. BIRDIE, RALPH F.; DRESSSEL, PAUL; and KELSO, PAUL. "Relative Validity of the Q and L Scores of the ACE Psychological Examination." *Educational and Psychological Measurement* 11: 803-12; Winter 1951.
30. BERK, ROBERT L. "Coaching in an Institution for Defective Delinquents: An Evaluation by Means of the Critical Incident Technique." *American Journal of Mental Deficiency* 56: 615-21; January 1952.
31. BERLINSKY, STANLEY. "Measurement of the Intelligence and Personality of the Deaf: A Review of the Literature." *Journal of Speech and Hearing Disorders* 17: 39-54; March 1952.
32. BESSENT, TRENT E. "A Note on the Validity of the Leiter International Performance Scale." *Journal of Consulting Psychology* 14: 234; June 1950.
33. BIRCH, JACK W. "The Goodenough Drawing Test and Older Mentally Retarded Children." *American Journal of Mental Deficiency* 54: 218-24; October 1949.
34. BIRCH, JANE R., and BIRCH, JACK W. "The Leiter International Performance Scale as an Aid in the Psychological Study of Deaf Children." *American Annals of the Deaf* 96: 502-11; November 1951.

35. BLAKE, ROBERT R. "The Relation Between Childhood Environment and the Scholastic Aptitude and Intelligence of Adults." *Journal of Social Psychology* 29: 37-41; February 1949.
36. BLOCK, JACK; LEVINE, LOUIS; and McNEMAR, QUINN. "Testing for the Existence of Psychometric Patterns." *Journal of Abnormal and Social Psychology* 46: 356-59; July 1951.
37. BOLTON, EURL. "Predictive Value for Academic Achievement of the ACE Psychological Examination Scores." *Peabody Journal of Education* 29: 345-60; May 1952.
38. BOND, GUY L., and FAY, LEO C. "A Comparison of the Performance of Good and Poor Readers on the Individual Items of the Stanford-Binet Scale, Forms L and M." *Journal of Educational Research* 43: 475-79; February 1950.
39. BORG, WALTER R. "A Study of the Relationship Between General Intelligence and Success in an Art College." *Journal of Educational Psychology* 40: 434-40; November 1949.
40. BRANDT, HYMAN. "Development and Construction of an Armed Services Qualification Test: I. Rationale, Item Content, and Construction." (Abstract) *American Psychologist* 4: 239; July 1949.
41. BRANDT, HYMAN, and BURKE, LAVERNE K. "Standardization of the Armed Forces Qualification Test AFQT-1 and 2." (Abstract) *American Psychologist* 5: 285; July 1950.
42. BROWN, HUGH S. "Differential Prediction by the ACE." *Journal of Educational Research* 44: 116-21; October 1950.
43. BURIK, THEODORE E. "Relative Roles of the Learning and Motor Factors Involved in the Digit Symbol Test." *Journal of Psychology* 30: 33-42; July 1950.
44. BURNHAM, CATHARINE A. "A Study of the Degree of Relationship Between Rorschach H% and Wechsler-Bellevue Arrangement Scores." *Rorschach Research Exchange* 13: 206-209; June 1949.
45. BUROS, OSCAR K., chairman. *Proceedings of the 1949 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1950. 94 p.
46. BURT, CYRIL. "Critical Notice: The Trend of Scottish Intelligence." *British Journal of Educational Psychology* 20: 55-61; February 1950.
47. BURT, CYRIL. "The Trend of National Intelligence." *British Journal of Sociology* 1: 154-68; June 1950.
48. BURTON, ARTHUR. "The Use of Psychometric and Projective Tests in Clinical Psychology." *Journal of Psychology* 28: 451-56; October 1949.
49. CANE, V. R., and HEIM, ALICE W. "The Effects of Repeated Retesting: III. Further Experiments and General Conclusions." *Quarterly Journal of Experimental Psychology* 2: 182-97; November 1950.
50. CARLSON, HILDING B., and HENDERSON, NORMAN. "The Intelligence of American Children of Mexican Parentage." *Journal of Abnormal and Social Psychology* 45: 544-51; July 1950.
51. CARRILLO, LAWRENCE W., JR., and REICHART, ROBERT R. "Use of a Caution Factor To Increase the Predictive Value of the ACE Examination for Students of Engineering." *Journal of Educational Research* 45: 361-68; January 1952.
52. CASSEL, ROBERT H. "Qualitative Evaluation of the Progressive Matrices Test." *Educational and Psychological Measurement* 9: 233-41; Summer 1949.
53. CASSEL, ROBERT H. "'A Rigorous Criterion of Feeble-mindedness': A Critique." *Journal of Abnormal and Social Psychology* 46: 116-17; January 1951.
54. CATTELL, RAYMOND B. "Classical and Standard Score IQ Standardization of the I.P.A.T. Culture-Free Intelligence Scale 2." *Journal of Consulting Psychology* 15: 154-59; April 1951.
55. CATTELL, RAYMOND B. "The Fate of National Intelligence; Test of a Thirteen-Year Prediction." *Eugenics Review* 42: 136-48; December 1950.
56. CHAUNCEY, HENRY. "The Use of the Selective Service College Qualification Test in the Deferment of College Students." *Science* 116: 73-79; July 25, 1952.
57. CHESLER, DAVID J. "The Wonderlic Personnel Test as a Predictor of Scores on the American Council on Education Examination." *Journal of Clinical Psychology* 4: 82-85; January 1948.
58. COCHRAN, SAMUEL W., and DAVIS, FREDERICK B. "Predicting Freshman Grades at George Peabody College for Teachers." *Peabody Journal of Education* 27: 352-56; May 1950.
59. COHEN, BERTRAM D., and COLLIER, MARY J. "A Note on the WISC and Other

- Tests of Children Six to Eight Years Old." *Journal of Consulting Psychology* 16: 226-27; June 1952.
60. COHEN, JACOB. "A Factor-Analytically Based Rationale for the Wechsler-Bellevue Subtests." *Journal of Consulting Psychology* 16: 272-77; August 1952.
  61. COHEN, JACOB. "Factors Underlying Wechsler-Bellevue Performance of Three Neuropsychiatric Groups." *Journal of Abnormal and Social Psychology* 47: 359-65; April 1952. (Supplement)
  62. COMBS, ARTHUR W. "Intelligence from a Perceptual Point of View." *Journal of Abnormal and Social Psychology* 47: 662-73; July 1952.
  63. COPPINGER, NEIL W., and AMMONS, ROBERT B. "The Full-Range Picture Vocabulary Test: VIII. A Normative Study of Negro Children." *Journal of Clinical Psychology* 8: 136-40; April 1952.
  64. CORNELL, ETHEL L., and GILLETTE, ANNETTE. "Construction and Educational Significance of Intelligence Tests." *Review of Educational Research* 20: 17-26; February 1950.
  65. CORSINI, RAYMOND. "The Immediate Test." *Journal of Clinical Psychology* 7: 127-30; April 1951.
  66. COTZIN, MILTON, and GALLACHER, JAMES J. "The Southbury Scale: A Valid Abbreviated Wechsler-Bellevue for Mental Defectives." *Journal of Consulting Psychology* 14: 358-64; October 1950.
  67. COZAN, LEE W. "Industrial Use of the Partington Pathways Test." *Journal of Applied Psychology* 35: 112-13; April 1951.
  68. CRONBACH, LEE J. *Essentials of Psychological Testing*. New York: Harper and Brothers, 1949. 475 p.
  69. CRUICKSHANK, WILLIAM M., and QUALTERE, THOMAS J. "The Use of Intelligence Tests with Children of Retarded Mental Development: I. Comparison of the 1916 and 1937 Revisions of the Stanford-Binet Intelligence Scales. II. Clinical Considerations." *American Journal of Mental Deficiency* 54: 361-69; 370-81; January 1950.
  70. CURTIS, HAZEN A. "A Study of the Relative Effects of Age and of Test Difficulty upon Factor Patterns." *Genetic Psychology Monographs* 40: 99-148; August 1949.
  71. DARCY, NATALIE T. "The Performance of Bilingual Puerto Rican Children on Verbal and on Non-Language Tests of Intelligence." *Journal of Educational Research* 45: 499-506; March 1952.
  72. DARLEY, JOHN G. "Review: Intelligence and Cultural Differences." *Journal of Applied Psychology* 36: 141-43; April 1952.
  73. DAVENPORT, K. S., and REMMERS, HERMANN H. "Factors in State Characteristics Related to Average A-12 V-12 Test Scores." *Journal of Educational Psychology* 41: 110-15; February 1950.
  74. DAVIDSON, KENNETH S., and OTHERS. "A Preliminary Study of Negro and White Differences on Form I of the Wechsler-Bellevue Scale." *Journal of Consulting Psychology* 14: 489-92; December 1950.
  75. DAVIS, ALLISON, and EELLS, KENNETH. *Davis-Eells Games*. Yonkers, N. Y.: World Book Co., 1952.
  76. DAVIS, ALLISON, and HESS, ROBERT D. "What about IQ's?" *Journal of the National Education Association* 38: 604-605; November 1949.
  77. DAVIS, PAUL C. "A Factor Analysis of the Wechsler-Bellevue Intelligence Scale, Form I, in a Matrix with Reference Variables." (Abstract) *American Psychologist* 7: 296-97; July 1952.
  78. DE GROOT, A. D. "The Effects of War upon the Intelligence of Youth." *Journal of Abnormal and Social Psychology* 43: 311-17; June 1948.
  79. DELATTRE, LOIS, and COLE, DAVID. "A Comparison of the WISC and the Wechsler-Bellevue." *Journal of Consulting Psychology* 16: 228-30; June 1952.
  80. DOPPELT, JEROME E. "Difficulty and Validity of Analogies Items in Relation to Major Field of Study." *Journal of Applied Psychology* 35: 30-33; February 1951.
  81. DOPPELT, JEROME E. *The Organization of Mental Abilities in the Age Range 13 to 17*. Contributions to Education, No. 962. New York: Teachers College, Columbia University, 1950. 86 p.
  82. DYER, HENRY S., chairman. *Proceedings of the 1951 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1952. 119 p.

83. DYER, HENRY S. "The Scholastic Aptitude Test—Items, Scores, and Coaching." *College Board Review* 15: 235-39; November 1951.
84. EDUCATIONAL TESTING SERVICE. *A Summary of Statistics on Selective Service College Qualification Test*. Princeton, N. J.: Educational Testing Service, 1952. 71 p.
85. ELLS, KENNETH, and OTHERS. *Intelligence and Cultural Differences*. Chicago: University of Chicago Press, 1951. 388 p.
86. EGLASH, ALBERT. "Validation of the Wechsler 'Shoes' Item." *Journal of Abnormal and Social Psychology* 45: 733-34; October 1950.
87. ELONEN, ANNA S. *A Comparison of Two Tests of Intelligence Administered to Adults*. Psychological Monographs, No. 306. Stanford, Calif.: Stanford University Press, 1949. 35 p.
88. ELWOOD, MARY I. "Changes in Stanford-Binet IQ of Retarded Six-Year-Olds." *Journal of Consulting Psychology* 16: 217-19; June 1952.
89. ENGLE, THELBURN L., and HAMLETT, IONA C. "Constancy of the IQ with Mentally Deficient Patients as Measured by the Time Appreciation Test." *American Journal of Mental Deficiency* 56: 775-76; April 1952.
90. ENGLE, THELBURN L., and HAMLETT, IONA C. "The Use of the Time Appreciation Test as a Screening or Supplementary Test for Mentally Deficient Patients." *American Journal of Mental Deficiency* 54: 521-25; April 1950.
91. ESCALONA, SIBYLLE. "The Use of Infant Tests for Predictive Purposes." *Bulletin of the Menninger Clinic* 14: 117-28; July 1950.
92. FINCH, FRANK H.; KUHLMANN, FREDERICK; and BETTS, GILBERT L. *Kuhlmann-Finch Intelligence Tests*. Philadelphia: Educational Test Bureau, 1952.
93. FINDLEY, WARREN G. "The Selective Service College Qualification Test." *American Psychologist* 6: 181-83; May 1951.
94. FINKELSTEIN, MELVILLE; GERBOTH, RENATE; and WESTERHOLD, RUTH. "Standardization of a Short Form of the Wechsler Vocabulary Subtest." *Journal of Clinical Psychology* 8: 133-35; April 1952.
95. FOULDS, G. A., and RAVEN, JOHN C. "An Experimental Survey with Progressive Matrices (1947)." *British Journal of Educational Psychology* 20: 104-10; June 1950.
96. FOX, CHARLOTTE, and BIRREN, JAMES E. "The Differential Decline of Subtest Scores of the Wechsler-Bellevue Intelligence Scale in 60-69-Year-Old Individuals." *Journal of Genetic Psychology* 77: 313-17; December 1950.
97. FRANDSEN, ARDEN N. "The Wechsler-Bellevue Intelligence Scale and High School Achievement." *Journal of Applied Psychology* 34: 406-11; December 1950.
98. FRANDSEN, ARDEN N., and HIGGINSON, JAY B. "The Stanford-Binet and the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 15: 236-38; June 1951.
99. FRANDSEN, ARDEN N.; McCULLOUGH, BETSEY R.; and STONE, DAVID R. "Serial Versus Consecutive Order Administration of the Stanford-Binet Intelligence Scales." *Journal of Consulting Psychology* 14: 316-20; August 1950.
100. FREDRIKSEN, NORMAN, and SCHRADER, WILLIAM B. "The ACE Psychological Examination and High School Standing as Predictors of College Success." *Journal of Applied Psychology* 36: 261-65; August 1952.
101. FREEMAN, FRANK S. *Theory and Practice of Psychological Testing*. New York: Henry Holt and Co., 1950. 518 p.
102. FRENCH, ELIZABETH G., and HUNT, WILLIAM A. "The Relationship of Scatter in Test Performance to Intelligence Level." *Journal of Clinical Psychology* 7: 95-98; January 1951.
103. FULK, BYRON E., and HARRELL, THOMAS W. "Negro-White Army Test Scores and Last School Grade." *Journal of Applied Psychology* 36: 34-35; February 1952.
104. FURST, EDWARD J. "Relationship Between Tests of Intelligence and Tests of Critical Thinking and of Knowledge." *Journal of Educational Research* 43: 614-25; April 1950.
105. FUSFELD, IRVING S. "On the ACE Psychological Examination." *School and Society* 70: 117-18; August 20, 1949.
106. GARRETT, HARLEY F. "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Science and Teachers Colleges." *Journal of Experimental Education* 18: 91-138; December 1949.
107. GELLERMAN, SAUL W. "Forms I and II of the Arthur Performance Scales with Mental Defectives." *Journal of Consulting Psychology* 16: 127-31; April 1952.

108. GELLERMAN, SAUL W., and HAYS, WILLIAM. "A Proposed Correction for the Confounded Effects of Cultural Variation in Intelligence Quotients." *American Journal of Mental Deficiency* 56: 177-79; July 1951.
109. GERBOTH, RENATE. "A Study of the Two Forms of the Wechsler-Bellevue Intelligence Scale." *Journal of Consulting Psychology* 14: 365-70; October 1950.
110. GILLILAND, ADAM R. "Environmental Influences on Infant Intelligence Test Scores." *Harvard Educational Review* 19: 142-46; Summer 1949.
111. GILLILAND, ADAM R. *The Northwestern Intelligence Tests: Test A. Test for Infants 4-12 Weeks Old; Test B. Test for Infants 13-36 Weeks Old*. Boston: Houghton Mifflin Co., 1949.
112. GILLILAND, ADAM R. "Socio-Economic Status and Race as Factors in Infant Intelligence Test Scores." *Child Development* 22: 271-73; December 1951.
113. GLASER, ROBERT. "The Validity of Some Tests for Predicting Achievement in Medical School." (Abstract) *American Psychologist* 6: 298; July 1951.
114. GOODENOUGH, FLORENCE L. *Mental Testing*. New York: Rinehart and Co., 1949. 609 p.
115. GOOSSEN, CARL V. "The Goossen Hidden Intelligence Test." *Public Opinion Quarterly* 14: 759-66; Winter 1950.
116. GORDON, LEONARD V., and DUREA, MERVIN A. "The Effect of Discouragement on the Revised Stanford-Binet Scale." *Journal of Genetic Psychology* 73: 201-207; December 1948.
117. GRAHAM, E. ELLIS. "Wechsler-Bellevue and WISC Scattergrams of Unsuccessful Readers." *Journal of Consulting Psychology* 16: 268-71; August 1952.
118. GROVE, WILLIAM B. "Mental Age Scores for the Wechsler Intelligence Scale for Children." *Journal of Clinical Psychology* 6: 393-97; October 1950.
119. GULLIKSEN, HAROLD. *Theory of Mental Tests*. New York: John Wiley and Sons, 1950. 486 p.
120. GURVITZ, MILTON S. "On the Decline of Performance on Intelligence Tests with Age." (Abstract) *American Psychologist* 6: 295; July 1951.
121. GURVITZ, MILTON S. "The Hillside Short Form of the Wechsler Bellevue." *Journal of Clinical Psychology* 7: 131-34; April 1951.
122. GURVITZ, MILTON S. "Some Defects of the Wechsler-Bellevue." *Journal of Consulting Psychology* 16: 124-26; April 1952.
123. GURVITZ, MILTON S. "Speed as a Factor in the Decline of Performance with Age." (Abstract) *American Psychologist* 7: 298-99; July 1952.
124. GURVITZ, MILTON S. "What Do Paper Formboards Measure?" (Abstract) *American Psychologist* 5: 278-79; July 1950.
125. HAGEN, ELIZABETH P. "A Factor Analysis of the Wechsler Intelligence Scale for Children." (Abstract) *American Psychologist* 6: 297; July 1951.
126. HAMILTON, MILDRED E. "A Comparison of the Revised Arthur Performance Tests (Form II) and the 1937 Binet." *Journal of Consulting Psychology* 13: 44-49; February 1949.
127. HANNA, JOSEPH V. "Estimating Intelligence by Interview." *Educational and Psychological Measurement* 10: 420-30; Autumn 1950.
128. HARPER, BERTHA P.; UHLANER, JULIUS E.; and MOSIER, CHARLES I. "Development and Construction of an Armed Services Qualification Test: II. Item Analysis and Item Selection." (Abstract) *American Psychologist* 4: 239-40; July 1949.
129. HAYES, SAMUEL P. "Measuring the Intelligence of the Blind." *Blindness: Modern Approaches to the Unseen Environment*. (Edited by Paul A. Zahl.) Princeton, N. J.: Princeton University Press, 1950. Chapter 10, p. 141-73.
130. HAYES, SAMUEL P. "Measuring the Intelligence of the Blind." *Psychological Diagnosis and Counseling of the Adult Blind*. (Edited by Wilma T. Donahue, and Donald H. Dabelstein.) New York: American Foundation for the Blind, 1950. p. 77-96.
131. HAYS, WILLIAM, and SCHNEIDER, BERNARD. "A Test-Retest Evaluation of the Wechsler Forms I and II with Mental Defectives." *Journal of Clinical Psychology* 7: 140-43; April 1951.
132. HEIL, WALTER G., and HORN, ALICE M. *A Comparative Study of the Data for Five Different Intelligence Tests Administered to 284 Twelfth Grade Students at South Gate High School, Los Angeles*. Los Angeles: Los Angeles City School Districts, Curriculum Division, February 1950. 25 p. (Mimeo.)

133. HELMICK, JOHN S. "Reliability or Variability." *Journal of Consulting Psychology* 16: 154-55; April 1952.
134. HERRING, FRED H. "An Evaluation of Published Short Forms of the Wechsler-Bellevue Scale." *Journal of Consulting Psychology* 16: 119-23; April 1952.
135. HICK, W. E. "Information Theory and Intelligence Tests." *British Journal of Psychology, Statistical Section* 4: 157-64; November 1951.
136. HILDEN, ARNOLD H., and TAYLOR, JAMES W. "Empirical Evaluation of Short W-B Scales." *Journal of Clinical Psychology* 8: 323-31; October 1952.
137. HOLDEN, RAYMOND H. "Improved Methods in Testing Cerebral Palsied Children." *American Journal of Mental Deficiency* 56: 349-53; October 1951.
138. HOLZBERG, JULES D., and BELMONT, LILLIAN. "The Relationship Between Factors on the Wechsler-Bellevue and Rorschach Having Common Psychological Rationale." *Journal of Consulting Psychology* 16: 23-29; February 1952.
139. HOLZINGER, KARL J., and CROWDER, NORMAN A. *Holzinger-Crowder Uni-Factor Tests*. Yonkers, N. Y.: World Book Co., 1952.
140. HUNT, WILLIAM A., and FRENCH, ELIZABETH G. "The Navy-Northwestern Matrices Test." *Journal of Clinical Psychology* 8: 65-74; January 1952.
141. HUNT, WILLIAM A., and FRENCH, ELIZABETH G. "The CVS Abbreviated Individual Intelligence Scale." *Journal of Consulting Psychology* 16: 181-86; June 1952.
142. INGHAM, J. G. "Memory and Intelligence." *British Journal of Psychology, General Section* 43: 20-31; February 1952.
143. JASTAK, JOSEPH. "Psychological Tests, Intelligence, and Feeble-mindedness." *Journal of Clinical Psychology* 8: 107-12; April 1952.
144. JASTAK, JOSEPH. "A Rigorous Criterion of Feeble-mindedness." *Journal of Abnormal and Social Psychology* 44: 367-78; July 1949.
145. JASTAK, JOSEPH. "On Robert H. Cassel's Critique of 'A Rigorous Criterion of Feeble-mindedness.'" *Journal of Abnormal and Social Psychology* 46: 118-19; January 1951.
146. JEWELL, BRUCE T., and WURSTEN, HELMUT. "Observations on the Psychological Testing of Cerebral Palsied Children." *American Journal of Mental Deficiency* 56: 630-37; January 1952.
147. JOHNSON, ELIZABETH Z. "Sex Differences and Variability in the Performance of Retarded Children on Raven, Binet and Arthur Tests." *Journal of Clinical Psychology* 8: 298-301; July 1952.
148. JONES, LYLE V. "A Factor Analysis of the Stanford-Binet at Four Age Levels." *Psychometrika* 14: 299-331; December 1949.
149. *Journal of Heredity*. "The Score of the Colleges." 43: 133-40; May-June 1952.
150. KEIR, GERTRUDE. "The Progressive Matrices as Applied to School Children." *British Journal of Psychology, Statistical Section* 2: 140-50; November 1949.
151. KELLY, EVERETT L., and FISKE, DONALD W. *The Prediction of Performance in Clinical Psychology*. Ann Arbor: University of Michigan Press, 1951. 311 p.
152. KENNEY, JOHN J. "Mathematics in Group Intelligence Tests." *Journal of Educational Research* 44: 129-33; October 1950.
153. KENT, GRACE H. *Mental Tests in Clinics for Children*. New York: D. Van Nostrand, 1950. 180 p.
154. KITZINGER, HELEN, and BLUMBERG, EUGENE. "Supplementary Guide for Administering and Scoring the Wechsler-Bellevue Intelligence Scale (Form I)." *Psychological Monographs*. Washington, D. C.: American Psychological Association, 1951. 20 p.
155. KNEHR, CHARLES A. "Intelligence As Structural Limitation and Potential." *Journal of Psychology* 29: 165-71; January 1950.
156. KNEHR, CHARLES A., and SOBOLEW, ALBERT. "Mental Ability of Prematurely Born Children at Early School Age." *Journal of Psychology* 27: 355-61; April 1949.
157. KNOTT, JOHN R., and OTHERS. "Brief Tests of Intelligence in the Psychiatric Clinic." *Journal of Clinical Psychology* 7: 123-26; April 1951.
158. KRUGMAN, JUDITH L., and OTHERS. "Pupil Functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 15: 475-83; December 1951.
159. KUHLMANN, FREDERICK, and ANDERSON, ROSE G. *Kuhlmann-Anderson Intelligence Test*. Sixth edition. Princeton, N. J.: Personnel Press, 1952.
160. LANNHOLM, GERALD V., and SCHRADER, WILLIAM B. *Predicting Graduate School Success*. Princeton, N. J.: Educational Testing Service, 1951. 50 p.
161. LEHMAN, CHARLES F. "An Investigation of Musical Achievement and Relationship to Intelligence and Musical Talent." *Journal of Educational Research* 45: 623-29; April 1952.

162. LEITER, RUSSELL G. "Caucasian Norms for the Leiter International Performance Scale." *Psychological Service Center Journal* 1: 136-38; December 1949.
163. LEITER, RUSSELL G. "The Leiter Adaptation of Arthur's Stencil Design Test." *Psychological Service Center Journal* 1: 62-68; September 1949.
164. LEITER, RUSSELL G. "The Leiter Adaptation of the Painted Cube Test." *Psychological Service Center Journal* 1: 29-45; September 1949.
165. LEITER, RUSSELL G. *Part II of the Manual for the 1948 Revision of the Leiter International Performance Scale*. Washington, D. C.: Psychological Service Center Press, 1952. 85 p.
166. LEITER, RUSSELL G., and PARTINGTON, JOHN E. *Leiter-Partington Adult Performance Scale*. Washington, D. C.: Psychological Service Center Press, 1950.
167. LENNON, ROGER T. "A Comparison of Results of Three Intelligence Tests." *Test Service Notebook No. 11*. Yonkers, N. Y.: World Book Co., 1951. 4 p.
168. LEVINE, ABRAHAM S. "Construction and Use of Verbal Analogy Items." *Journal of Applied Psychology* 34: 105-107; April 1950.
169. LEVINE, ABRAHAM S. "Correcting Special Ability Test Scores for General Ability." *Journal of Applied Psychology* 33: 566-68; December 1949.
170. LEVINE, ABRAHAM S. "Minnesota Psycho-Analogies Test." *Journal of Applied Psychology* 34: 300-305; October 1950.
171. LINDZEY, GARDNER. "Remarks on the Use of the Wonderlic Personnel Test as a 'Pre-test.'" *Journal of Clinical Psychology* 5: 100-102; January 1949.
172. LORGE, IRVING, and KRUGLOV, LORRAINE. "The Relation Between Merit of Written Expression and Intelligence." *Journal of Educational Research* 44: 507-19; March 1951.
173. LORGE, IRVING, and KRUGLOV, LORRAINE. "The Relationship Between the Readability of Pupils' Compositions and Their Measured Intelligence." *Journal of Educational Research* 43: 467-74; February 1950.
174. MCKEE, JOHN P. "The Tests of Primary Mental Abilities Applied to Superior Children." *Journal of Educational Psychology* 43: 45-56; January 1952.
175. MCNEMAR, QUINN. "On Abbreviated Wechsler-Bellevue Scales." *Journal of Consulting Psychology* 14: 79-81; April 1950.
176. MCNEMAR, QUINN. "Review: Intelligence and Cultural Differences." *Psychological Bulletin* 49: 370-71; July 1952.
177. MACARET, ANN, and THOMPSON, CLARE W. "Differential Test Responses of Normal, Superior and Mentally Defective Subjects." *Journal of Abnormal and Social Psychology* 45: 163-67; January 1950.
178. MANOLAKES, GEORGE, and SHELDON, WILLIAM D. "A Comparison of the Grace Arthur, Revised Form II, and the Stanford-Binet, Revised Form L." *Educational and Psychological Measurement* 12: 105-108; Spring 1952.
179. MANUEL, HERSCHEL T. *Cooperative Inter-American Tests: General Ability, Reading, Social Studies, Natural Sciences, Language Usage*. Princeton, N. J.: Educational Testing Service, 1950.
180. MANUEL, HERSCHEL T. "The Inter-American Series of Parallel Tests for Children of Different Languages." *American Journal of Mental Deficiency* 54: 93-100; July 1949.
181. MANUEL, HERSCHEL T. "The Use of Tests in Latin American Countries." *Journal of Educational Research* 44: 529-33; March 1951.
182. MATARAZZO, JOSEPH D. "A Study of the Diagnostic Possibilities of the CVS with a Group of Organic Cases." *Journal of Clinical Psychology* 6: 337-43; October 1950.
183. MATTHEWS, JACK, and BIRCH, JACK W. "The Leiter International Performance Scale—a Suggested Instrument for Psychological Testing of Speech and Hearing Clinic Cases." *Journal of Speech and Hearing Disorders* 14: 318-21; December 1949.
184. MEISTER, RALPH K., and KURKO, VIRGINIA K. "An Evaluation of a Short Administration of the Revised Stanford-Binet Intelligence Examination." *Educational and Psychological Measurement* 11: 489-93; Autumn 1951.
185. MENSCH, IVAN N. "Brief Psychological Measures." *Nervous Child* 8: 349-59; July 1949.
186. MILLARD, KENNETH A. "Is How Supervise? an Intelligence Test?" *Journal of Applied Psychology* 36: 221-24; August 1952.
187. MILLER, WILFORD S. *Miller Analogies Test*. New York: Psychological Corp., 1947.
188. MURSELL, GEORGE R. "A Simplified Case (Box) for the Administration of the

- Kuhlmann Scale of Mental Development." *American Journal of Mental Deficiency* 56: 791-95; April 1952.
189. NALE, STANLEY L. "The Childrens Wechsler and the Binet on 104 Mental Defectives at the Polk State School." *American Journal of Mental Deficiency* 56: 419-23; October 1951.
190. NEWTON, RICHARD L. "A Comparison of Two Methods of Administering the Digit Span Test." *Journal of Clinical Psychology* 6: 409-12; October 1950.
191. NOTCUTT, BERNARD. "The Distribution of Scores on Raven's Progressive Matrices Test." *British Journal of Psychology, General Section* 40: 68-70; December 1949.
192. OSBORNE, R. TRAVIS; SANDERS, WILMA B.; and GREENE, JAMES E. "The Differential Prediction of College Marks by ACE Scores." *Journal of Educational Research* 44: 107-15; October 1950.
193. OTIS, ARTHUR S. *Otis Quick-Scoring Mental Ability Test: New Edition, Alpha Test, Form A-s (Short Form)*. Yonkers, N. Y.: World Book Co., 1952.
194. OTIS, JAY L., and CHESLER, DAVIS J. "A Short Test of Mental Ability." *Journal of Applied Psychology* 33: 146-50; April 1949.
195. PARTINGTON, JOHN E. "Detailed Instructions for Administering Partington's Pathways Test." *Psychological Service Center Journal* 1: 46-48; September 1949.
196. PARTINGTON, JOHN E., and LEITER, RUSSELL G. "Partington's Pathways Test." *Psychological Service Center Bulletin* 1: 9-20; March 1949.
197. PASCAL, GERALD R., and OTHERS. "The Delayed Reaction in Mental Defectives." *American Journal of Mental Deficiency* 56: 152-60; July 1951.
198. PASTORE, NICHOLAS. "A Fallacy Underlying Garrett's Use of the Data of the Army Alpha and Beta Tests—A Comment." *Scientific Monthly* 69: 279-80; October 1949.
199. PASTOVIC, JOHN J., and GUTHRIE, GEORGE M. "Some Evidence on the Validity of the WISC." *Journal of Consulting Psychology* 15: 385-86; October 1951.
200. PEARSON, JOHN S., and STRATE, MARVIN W. "The Minnesota Psycho-Analogies Test in the Selection of Psychologists for Public Service." *Journal of Applied Psychology* 35: 314-15; October 1951.
201. PEEL, EDWIN A. "A Note on Practice Effects in Intelligence Tests." *British Journal of Educational Psychology* 21: 122-25; June 1951.
202. PENROSE, LIONEL S. "Genetical Influences on the Intelligence Level of the Population." *British Journal of Psychology, General Section* 40: 128-36; March 1950.
203. PENROSE, LIONEL S. "Propagation of the Unfit." *Lancet* 259: 425-27; September 1950.
204. Personnel Research Section, Adjutant General's Office. *PRS Report 778: Comparison of Army and Navy Classification Tests*. Washington, D. C.: the Section, 1949. 18 p.
205. PIERCE, HELEN O. "Errors Which Can and Should Be Avoided in Scoring the Stanford-Binet Scale." *Journal of Genetic Psychology* 72: 303-305; June 1948.
206. PIERCE-JONES, JOHN, and TYLER, FRED T. "A Comparison of the A. C. E. Psychological Examination and the Culture-Free Test." *Canadian Journal of Psychology* 4: 109-14; September 1950.
207. POLLACZEK, PENELOPE P. "A Study of Malingering on the CVS Abbreviated Individual Intelligence Scale." *Journal of Clinical Psychology* 8: 75-81; January 1952.
208. PORTEUS, STANLEY D. *The Porteus Maze Test and Intelligence*. Palo Alto, Calif.: Pacific Books, 1950. 194 p.
209. PORTEUS, STANLEY D. "Recent Research on the Porteus Maze Test and Psycho-Surgery." *British Journal of Medical Psychology* 24: 132-40; June 1951.
210. PORTEUS, STANLEY D. "Thirty-Five Years' Experience with the Porteus Maze." *Journal of Abnormal and Social Psychology* 45: 396-401; April 1950.
211. PRESSEY, SIDNEY L. *Educational Acceleration: Appraisals and Basic Problems*. Bureau of Educational Research Monographs, No. 31. Columbus: Ohio State University, 1949. 153 p.
212. RABIN, ALBERT I., and GUERTIN, WILSON H. "Research with the Wechsler-Bellevue Test: 1945-1950." *Psychological Bulletin* 48: 211-48; May 1951.
213. RAMASESHAN, RUKMINI S. "A Note on the Validity of the Mental Age Concept." *Journal of Educational Psychology* 41: 56-58; January 1950.
214. RAVEN, JOHN C. "The Instinctive Disposition To Act Intelligently." *British Journal of Psychology, General Section* 42: 336-44; November 1951.

215. RICHARDS, THOMAS W. "Mental Test Performance as a Reflection of the Child's Current Life Situation: A Methodological Study." *Child Development* 22: 221-33; September 1951.
216. RIMOLDI, H. J. A. "The Central Intellective Factor." *Psychometrika* 16: 75-101; March 1951.
217. ROBERTS, J. A. FRASER, and MELLONE, MARGARET A. "On the Adjustment of Terman-Merrill I.Q.'s To Secure Comparability at Different Ages." *British Journal of Psychology, Statistical Section* 5: 65-79; June 1952.
218. ROE, ANNE. "Psychological Tests of Research Scientists." *Journal of Consulting Psychology* 15: 492-95; December 1951.
219. ROSILDA, SISTER M. "Is an I.Q. an Index to Algebra Ability?" *Journal of Educational Research* 44: 391-93; January 1951.
220. RUDOLF, G. DE M. "Re-testing of the Intelligence Quotient and the Social Age." *Journal of Mental Science* 95: 696-702; July 1949.
221. SACKS, ELINOR L. "Intelligence Scores as a Function of Experimentally Established Social Relationship Between Child and Examiner." *Journal of Abnormal and Social Psychology* 47: 354-58; April 1952 (Supplement)
222. SANDERCOCK, M. G., and BUTLER, A. J. "An Analysis of the Performance of Mental Defectives on the Wechsler Intelligence Scale for Children." *American Journal of Mental Deficiency* 57: 100-105; July 1952.
223. SCHERER, ISIDOR W. "The Psychological Scores of Mental Patients in an Individual and Group Testing Situation." *Journal of Clinical Psychology* 5: 405-408; October 1949.
224. SEASHORE, HAROLD G. "Differences Between Verbal and Performance IQ's on the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 15: 62-67; February 1951.
225. SEASHORE, HAROLD G.; WESMAN, ALEXANDER G.; and DOPPELT, JEROME E. "The Standardization of the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 14: 99-110; April 1950.
226. SEMEONOFF, B., and LAIRD, A. J. "The Vigotsky Test as a Measure of Intelligence." *British Journal of Psychology* 43: 94-102; May 1952.
227. SHANNON, WALTER, and ROSSI, PHILIP D. "Suggestions for Efficient Presentation of the Wechsler-Bellevue Object-Assembly Sub-test." *Journal of Clinical Psychology* 8: 413-15; October 1952.
228. SHUEY, AUDREY M. "Improvement in Scores on the American Council Psychological Examination from Freshman to Senior Year." *Journal of Educational Psychology* 39: 417-26; November 1948.
229. SHUEY, AUDREY M., and HEDRICK, CARROLL M. "Intelligence of College Women as Related to Family Size." *Journal of Educational Psychology* 42: 215-22; April 1951.
230. SINHA, UMA. "A Study of the Reliability and Validity of the Progressive Matrices Test." (Abstract) *British Journal of Educational Psychology* 21: 238-39; November 1951.
231. SKODAK, MARIE. "Mental Growth of Adopted Children in the Same Family." *Journal of Genetic Psychology* 77: 3-9; September 1950.
232. SKODAK, MARIE, and SKEELS, HAROLD M. "A Final Follow-up Study of One Hundred Adopted Children." *Journal of Genetic Psychology* 75: 85-125; September 1949.
233. SLOAN, WILLIAM. "Motor Proficiency and Intelligence." *American Journal of Mental Deficiency* 55: 394-406; January 1951.
234. SLOAN, WILLIAM, and SCHNEIDER, BERNARD. "A Study of the Wechsler Intelligence Scale for Children with Mental Defectives." *American Journal of Mental Deficiency* 55: 573-75; April 1951.
235. SMITH, ARTHUR E. *A Comparison of the SRA Primary Mental Abilities Test with the Wechsler-Bellevue Intelligence Scale*. Master's thesis. Normal: Illinois State Normal University, 1949. 24 p. (Typewritten)
236. SMITH, HERBERT A. "The Relationship Between Intelligence and the Learning Which Results from the Use of Educational Sound Motion Pictures." *Journal of Educational Research* 43: 241-47; December 1949.
237. STACEY, CHALMERS L., and LEVIN, JANICE. "Correlation Analysis of Scores of Subnormal Subjects on the Stanford-Binet and Wechsler Intelligence Scale for Children." *American Journal of Mental Deficiency* 55: 590-97; April 1951.
238. STACEY, CHALMERS L., and MARKIN, KARL E. "A Study of the Differential Responses among Three Groups of Subnormals on the Similarities Sub-Test

- of the Wechsler Intelligence Scale." *American Journal of Mental Deficiency* 56: 424-28; October 1951.
239. STACEY, CHALMERS L., and PORTNOY, BERNARD. "A Study of the Differential Responses on the Vocabulary Sub-Test of the Wechsler-Bellevue Intelligence Scale." *Journal of Clinical Psychology* 7: 144-48; April 1951.
  240. STAFFORD, JOHN W. "The Prediction of Success in Graduate School." (Abstract) *American Psychologist* 6: 298; July 1951.
  241. STANLEY, JULIAN C. "Five Recent Educational and Psychological Measurement Textbooks." *Harvard Educational Review* 22: 57-61; Winter 1952.
  242. STANLEY, JULIAN C. "A Note on the Correlation Between Nonlanguage Mental Ages and Reading Test Scores." *Journal of the Tennessee Academy of Science* 26: 88, 92; January 1951.
  243. STEELE, JOHN E. "Tests Used in Recruiting and Selecting College Graduates." *Personnel*, 26: 200-204; November 1949.
  244. STEISEL, IRA M. "The Relation Between Test and Retest Scores on the Wechsler-Bellevue Scale (Form I) for Selected College Students." *Journal of Genetic Psychology* 79: 155-62; December 1951.
  245. STEISEL, IRA M. "Retest Changes in Wechsler-Bellevue Scores as a Function of the Time Interval Between Examinations." *Journal of Genetic Psychology* 79: 199-203; December 1951.
  246. STEPHENSON, WILLIAM. *Testing School Children*. New York: Longmans, Green and Co., 1949. 127 p.
  247. STONESIFER, FRED A. "A Goodenough Scale Evaluation of Human Figures Drawn by Schizophrenic and Non-Psychotic Adults." *Journal of Clinical Psychology* 5: 396-98; October 1949.
  248. STORRS, SIBYLL V. "Evaluative Data on the G.A.T.B." *Personnel and Guidance Journal* 31: 87-90; November 1952.
  249. SUPER, DONALD E. *Appraising Vocational Fitness*. New York: Harper and Brothers, 1949. 715 p.
  250. SWANSON, EDWARD O. "The Relation of Vocabulary Test-Retest Gains to Amount of College Attendance after a Twenty-Four Year Period." (Abstract) *American Psychologist* 7: 368; July 1952.
  251. SWINEFORD, FRANCES. "General, Verbal, and Spatial Bi-factors after Three Years." *Journal of Education Psychology* 40: 353-60; October 1949.
  252. SWINEFORD, FRANCES. "The Nature of the General, Verbal, and Spatial Bi-Factors." *Supplementary Educational Monographs* 67: 1-71; November 1948.
  253. TAMMINEN, A. W. "A Comparison of the Army General Classification Test and the Wechsler-Bellevue Intelligence Scales." *Educational and Psychological Measurement* 11: 646-55; Winter 1951.
  254. TATE, MIRIAM E. "The Influence of Cultural Factors on the Leiter International Performance Scale." *Journal of Abnormal and Social Psychology* 47: 497-501; April 1952. (Supplement)
  255. THOMSON, GODFREY H. "Intelligence and Fertility in Scotland." *Eugenical News* 35: 23-24; March-June 1950.
  256. THOMSON, GODFREY H. "Intelligence and Fertility; The Scottish 1947 Survey." *Eugenics Review* 41: 163-70; January 1950.
  257. THOMPSON, GODFREY H., chairman. *The Trend of Scottish Intelligence: A Comparison of the 1947 and 1932 Surveys of the Intelligence of Eleven-Year-Old Pupils*. London: University of London Press, 1949. 151 p.
  258. THORNDIKE, ROBERT L. "Community Variables as Predictors of Intelligence and Academic Achievement." *Journal of Educational Psychology* 42: 321-38; October 1951. "Note of Correction." 43: 179-80; March 1952.
  259. THURSTONE, THELMA G., and THURSTONE, LOUIS L. *Thurstone Test of Mental Alertness*. Chicago: Science Research Associates, 1952.
  260. TILTON, JOHN R. "A Survey of the Reliability, Validity, and Usefulness of the Cattell Culture-Free Test." *Persona* 1: 17-19; Summer-Fall 1949.
  261. TILTON, JOHN W. "Intelligence Test Scores as Indicative of Ability To Learn." *Educational and Psychological Measurement* 9: 291-96; Autumn 1949.
  262. TIZARD, J. "The Porteus Maze Test and Intelligence: A Critical Survey." *British Journal of Educational Psychology* 21: 172-85; November 1951.
  263. TRACHT, VERNON S. "Preliminary Findings on Testing the Cerebral Palsied with Raven's 'Progressive Matrices.'" *Journal of Exceptional Children* 15: 77-79, 89; December 1948.

264. TRAVERS, ROBERT M. W., and WALLACE, WIMBURN L. "The Assessment of the Academic Aptitude of the Graduate Student." *Educational and Psychological Measurement* 10: 371-79; Autumn 1950.
265. TRAVERS, ROBERT M. W., and WALLACE, WIMBURN L. "Inconsistency in the Predictive Value of a Battery of Tests." *Journal of Applied Psychology* 34: 237-39; August 1950.
266. TRAXLER, ARTHUR E. "Reliability and Validity of the Scores on the Six Parts of the American Council on Education Psychological Examination." *Educational Records Bulletin* 58: 71-79; March 1952.
267. TRAXLER, ARTHUR E. "Twelve Years of Experience with the Junior Scholastic Aptitude Test." *Educational Records Bulletin* 59: 79-92; July 1952.
268. VERNON, PHILIP E. "Psychological Studies of the Mental Quality of the Population." *British Journal of Educational Psychology* 20: 35-42; February 1950.
269. VERNON, PHILIP E. "Recent Investigations of Intelligence and Its Measurement." *Eugenics Review* 43: 125-37; December 1951.
270. VERNON, PHILIP E. *The Structure of Human Abilities*. New York: John Wiley and Sons, 1950. 160 p.
271. WALLACE, JEAN G. "Results of a Test of High-Grade Intelligence Applied to a University Population." *British Journal of Psychology, General Section* 43: 61-69; February 1952.
272. WALLACE, WIMBURN L. "Differential Predictive Value of the ACE Psychological Examination." *School and Society* 70: 23-25; July 9, 1949.
273. WALLACE, WIMBURN L. "The Prediction of Grades in Specific College Courses." *Journal of Educational Research* 44: 587-97; April 1951.
274. WEAVER, HERBERT B. "The Leiter-Partington Adult Performance Scale at College Level." *Psychological Service Center Journal* 2: 182-88; September 1950.
275. WEBB, WILSE B. "Corrections for Variability: A Reply." *Journal of Consulting Psychology* 16: 156; April 1952.
276. WEBB, WILSE B., and DE HAAN, HENRY. "Wechsler-Bellevue Split-Half Reliabilities in Normals and Schizophrenics." *Journal of Consulting Psychology* 15: 68-71; February 1951.
277. WEBB, WILSE B., and HANER, CHARLES. "Quantification of the Wechsler-Bellevue Vocabulary Sub-Test." *Educational and Psychological Measurement* 9: 693-707; Winter 1949.
278. WECHSLER, DAVID. "Cognitive, Conative, and Non-Intellective Intelligence." *American Psychologist* 5: 78-83; January 1950.
279. WECHSLER, DAVID. "Equivalent Test and Mental Ages for the WISC." *Journal of Consulting Psychology* 15: 381-84; October 1951.
280. WECHSLER, DAVID. "Intellectual Development and Psychological Maturity." *Child Development* 21: 45-50; March 1950.
281. WECHSLER, DAVID. *Wechsler Intelligence Scale for Children Manual*. New York: Psychological Corp., 1949. 113 p.
282. WEIDER, ARTHUR; NOLLER, PAUL A.; and SCHRAMM, THEODORE A. "The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet." *Journal of Consulting Psychology* 15: 330-33; August 1951.
283. WHEELER, LESTER R. "The Relation of Reading to Intelligence." *School and Society* 70: 225-27; October 1949.
284. WICKERT, FREDERICK R. "Relation Between How Supervise? Intelligence, and Education for a Group of Supervisory Candidates in Industry." *Journal of Applied Psychology* 36: 301-303; October 1952.
285. WILLIAMS, NANCY. "A Study of the Validity of the Verbal Reasoning Subtest and the Abstract Reasoning Subtest of the Differential Aptitude Tests." *Educational and Psychological Measurement* 12: 129-31; Spring 1952.
286. WITTENBORN, JOHN R., and HOLZBERG, JULES D. "The Wechsler-Bellevue and Descriptive Diagnosis." *Journal of Consulting Psychology* 15: 325-29; August 1951.
287. WOLFLE, DAEL. "The Human Resources of the U. S.: Intellectual Resources." *Scientific American* 185: 42-46; September 1951.
288. WOLFLE, DAEL, and OXToby, TOBY. "Distributions of Ability of Students Specializing in Different Fields." *Science* 116: 311-14; September 26, 1952.
289. YOUNG, FLORENE M., and PITTS, VIRGINIA A. "Performance of Congenital Syphilitics on the Wechsler Intelligence Scale for Children." *Journal of Consulting Psychology* 15: 239-42; June 1951.
290. ZAGORSKI, HENRY J. *A Pattern Analysis of the Miller Analogies Test*. Master's thesis. Pittsburgh: University of Pittsburgh, 1949. 55 p. (Typewritten)

## CHAPTER III

### Development and Applications of Tests of Special Aptitude

WILLIAM G. MOLLENKOPF

THE field of special aptitude tests has been an active one during the past three years. Not only have there been new tests, including one which serves as an instrument of national manpower policy, but also there have been numerous studies of the effectiveness of tests and considerable efforts to increase their effectiveness, both for prediction in a single field and for differentiating among fields. The attention given to theoretical and rational considerations of test validity, and especially to the problem of the criterion, is especially significant.

#### The Selective Service College Qualification Test

Of the new tests which appeared during the past three years, the one of greatest general significance was the *Selective Service College Qualification Test*. Findley (50) described the specifications and initial plans for this test. Designed as an educational aptitude test intended to give no special advantage to students of any particular field, it contained 150 items, with an equal emphasis on verbal and quantitative abilities. The four chief item types of the forms used in 1951 were reading comprehension, verbal relations, arithmetic reasoning, and interpretation of data. Items were included only after try-out and analysis and were arranged in spiral blocks of 15 or 30 items, graded in difficulty. While a time limit of three hours was employed, the test was primarily a power measure. The test was scaled against the *Army General Classification Test* used in World War II so that a score of 70 on SSCQT is comparable to an AGCT score of 120, whereas a score of 75 corresponds to an AGCT score of 130.

Chauncey (27) further described steps leading up to the development of SSCQT, and provided a summary of findings of studies of (a) regional differences in test performances and differences among students in various major fields, and (b) relationship between test performance and college rank-in-class.

Comparison of the percentages of applicants in various geographic regions revealed that the proportion of students from New England, Middle Atlantic, East North-Central, West North-Central, and Pacific regions who earned scores of 70 or higher was somewhat higher than for the country as a whole. The percentage passing the test was well above average for those whose major field was engineering or the physical sciences and mathematics, whereas the percentage at or above 70 was well below the average for students in business and commerce, agriculture, and education.

Data on class standing were obtained in advance of test administration

for 5527 students at 23 selected colleges and universities. Tremendous variability was observed among these institutions in score-level; for example, the percentage of liberal-arts freshmen who achieved a score of 70 or more varied from 35 to 98 in 14 different groups. Despite these wide differences among institutions, the variation among coefficients of correlation between test score and rank in class for the various groups varied no more (.41 to .74) than would be expected on the basis of sampling fluctuation. The test thus basically appeared to be as good a predictor of freshman grades at one institution as at another. For six freshman groups who took both SSCQT and the *College Board Scholastic Aptitude Test*, the average correlation with rank in class was .52 for SSCQT and .53 for SAT. For 13 groups of freshmen who took the SSCQT and also the *ACE Psychological Examination*, the average correlation with rank in class was .53 for SSCQT and .41 for ACEPE.

### Medical College Selection Tests

Several studies of the *Professional Aptitude Test*, the forerunner of the present widely-used *Medical College Admission Test*, appeared. Ralph and Taylor (133) carried out a study of 44 medical students at the University of Utah. Correlations of scores on parts of the *Professional Aptitude Test* with grades for the first five quarters ranged from  $-.06$  to  $+.26$ . These authors contrasted with the above findings the correlations for several *General Aptitude Test Battery* scores: .47 for G; .45 for V; .39 for N; and .41 for S. Another study of the PAT was reported by Glaser (64). Scores on various parts of the test for a group of 150 students at the Indiana University School of Medicine correlated from .22 to .39 with first-year grade-point average. In neither study was it clearly indicated to what extent PAT scores had been used in selection.

An example of the drastic effects of sharp selection on the size of validity coefficients was given by Morris (115). In his study of correlations between parts of the PAT and first-year grade-point average in medicine for 81 students at the State University of Iowa, the coefficients ranged from .17 to .48. When he corrected the .48 for restriction of range, it rose to .73.

In October 1948, the PAT was succeeded by the *Medical College Admission Test* (MCAT). Stalnaker (141) indicated that the purpose of the new test (the official test of the Association of American Medical Colleges) was to give each college an independent common index for all its applicants, to be used in selection in conjunction with other evidence. Taylor (145) checked the validity of the MCAT by correlating part scores with grade-point averages for 42 members of the class entering in 1948 and 45 entering in 1949 at Utah. For the 1948 group, correlations ranged from .02 to .30, and for the 1949 group, from  $-.16$  to  $+.31$ . However, there was evidence that selection had drastically reduced the range of talent: standard deviations for verbal ability were 70 and 67 in the two years, and those for premedical science were 66 and 66, whereas the standard deviation for unselected candidates is 100. Ralph and Taylor (134) commented that

for five samples of medical students from three universities, it was evident that these students were more highly selected on certain MCAT subtest characteristics than on others.

Schultz (137) in a study of the science test of the MCAT, involving candidates from five large private universities, found no support for the hypothesis that taking extra courses in biology, chemistry, or physics beyond a certain minimum level would lead to better scores on this test.

### Tests for Engineering

In 1949, Moore (114) reviewed the previous 10 years of research on the selection of engineering students. A number of tests were found especially effective: the *Engineering and Physical Science Aptitude Test*; the *College Entrance Examination Board Mathematics Test*; the *Iowa Mathematics Aptitude Test*; the *Iowa Chemistry Aptitude Test*; the *Iowa Physics Aptitude Test*; and the *Pre-Engineering Inventory*.

Lord, Cowles, and Cynamon (97) described the *Pre-Engineering Inventory*, and reported results of an extensive study of its validity in 12 engineering schools. Median correlation coefficients for the seven parts ranged from .35 to .58. The composite score, derived from the second, third, and fourth parts, yielded a median validity coefficient of .60.

Another study involving the *Pre-Engineering Inventory* was that of Pierson and Jex (127). For a group of 276 first-year engineering students at the University of Utah, various multiple correlations of combinations of *Inventory Tests* and high-school grade-point ratios with the criterion of first-year-college grade-point ratios were in the high .60's.

Johnson (81) indicated that while the *Pre-Engineering Inventory* was still to be available for administration by various institutions, it was last administered in a nationwide program in June 1949. However, in December 1949 a new test, the *Pre-Engineering Science Comprehension Test*, was added to the examinations offered during administrations of the *College Entrance Examination Board Tests*. Johnson also reported a correlation of .66 for a combination of M-scores on the *Scholastic Aptitude Test* and high-school grades with first-year engineering grades for a total of 721 freshmen at five universities; the validity of high-school grades alone was .46.

Using as their criterion the first-semester grade-point averages of 192 beginning engineering students, Treumann and Sullivan (155) found a validity of .53 for scores on the *Engineering and Physical Science Aptitude Test*. For a group composed of most of these students, high-school rank gave a correlation of .49 with grades. Gregg (68) reported a further study of the validity of the EPSAT based on a group of 344 male and 8 female engineering freshmen at the University of Colorado. Right scores on the test correlated .58 with a weighted sum of grades in five freshman courses; the correlation was .63 when the scores were corrected for guessing.

In the study by Berdie and Sutter (17) of 372 engineering students at the University of Minnesota, the most effective predictor was rank in

high-school graduating class. However, the tests used were different from those mentioned above, and in no case were they especially designed for prediction of success in engineering. Berdie (16), reporting on the effectiveness of the *Differential Aptitude Tests* as predictors in engineering, indicated that the tests were not appropriate in difficulty and range for use in predicting success in engineering training when given at the college level.

Mandell and Chad (104) described several studies in which tests were given to engineers in the federal government. A version of the *Gottschaldt Figures Test* prepared by L. L. Thurstone yielded biserial correlations of .59, .47, and .57 for predicting an upper-lower group criterion derived by dividing engineers at a given salary grade into two groups according to age and time in grade, the three groups being 36 engineers at the Naval Electronics Laboratory, 38 at Naval Air Materiel Center, and 55 at the Vicksburg Corps of Engineers District Office. In the same set of groups a formulation test and an abstract reasoning test yielded median validities in the 'forties. In another article Mandell (102) reported a correlation of .32 between spatial visualization scores and ratings of the job performances of 114 aeronautical and mechanical engineers. It must, of course, be noted that employed workers were tested; these were not predictive studies.

### Legal Aptitude Tests

In his survey of 27 law schools found to be using legal aptitude tests, Feeney (49) found four in use: the *Law School Admission Test*, the *Iowa Legal Aptitude Test*, the *Ferson-Stoddard Law Aptitude Examination*, and a test constructed by one school for its own use. Of the 27 schools, 17 were using the LSAT. In a study conducted at 12 law schools, the correlation of prelaw grades with first-year law grades was found to be .38, the corresponding validity of LSAT scores was .40, and that for a weighted composite was .52. Johnson (79) presented further validity data for the 12-law-school study, which involved a total of 1725 day students. His article is noteworthy in that it is one of the few instances in the literature in which an abac is provided; this one was for determining the most likely law-school grade, and the chances in 100 for exceeding any selected grade, thru use of the average prelaw grades and LSAT score.

An interesting description of how Yale Law School uses results from the *Law School Admission Test* was provided by Braden (19). Relative emphasis placed on college grades and LSAT score varied according to which of three groups the college was placed in, on the basis of studies of the goodness of its grades for predicting law-school success.

### Selection in Other Professional Fields

According to Peterson (125), beginning with the class entering dental school in the fall of 1951 applicants were to be asked to take a battery of examinations administered by the Council on Dental Education of the

American Dental Association. The battery was the outcome of a program of aptitude testing conducted by Peterson since 1946. One of the tests is a *Carving Dexterity Test*. The applicant is given 80 minutes to carve two patterns from two large pieces of chalk; scoring is based on accuracy of dimensions, cleanness of angles, symmetry, and flatness of surfaces. Weiss (159) reported a study of the validity of this test at the School of Medicine, University of Kansas; scores correlated from .24 to .35 with technic grades in the classes of 1946-1948, each numbering approximately 100.

Procedures for the improvement of selection of personnel for public accounting were described by Traxler (154). An aptitude measure, termed an *Orientation Test*, resulted from a project sponsored by the American Institute of Accountants. It yields a verbal score and a quantitative score. Validities against college grades in accounting were stated to be .33 for verbal, .44 for quantitative, and .43 for the total score. A median correlation of .35 was reported between test scores and supervisors' ratings.

An *Administrative-Judgment Test* designed to measure understanding of administrative problems of large organizations was described by Mandell (100). For 258 cases the split-half reliability was .94. When several small groups of persons in administrative work in the federal government were given the test, and scores were correlated against the criteria of ratings of job performance and of position grade or salary, the median of seven coefficients was .51, and six of the coefficients were significant at the 1 percent level.

An aptitude test designed to predict scholastic success in the first professional year of veterinary medicine was reported by Owens (122). Tetrachoric correlations between scores and grade-point averages at Cornell, Iowa State, Kansas State, and Michigan State ranged from .48 to .72.

Levine (92) developed an evaluation instrument in psychology which was termed the *Minnesota Psycho-Analogies Test*. Items followed the analogy form, the first part of each item containing general vocabulary and information, the second part being psychological in character. Pearson and Strate (124) found a rank-difference correlation of .56 between combined scores on Forms A and B of this test and a ranking of 23 psychologists employed in the Minnesota Civil Service Department.

Travers and Wallace (152) described a test built to predict graduate-school success at the University of Michigan. Validation studies were carried out on graduate students in five fields. A comparison of the multiple correlations obtained from parts of the new test with the validities of the *Miller Analogies Test* favored the new test, but these multiples required negative weights in several instances.

Mandell (103) indicated that, in a study in the federal government, scores on Engelhart's *Hypotheses Test* correlated .39, .44, and .41 with salary for three groups of chemists numbering respectively 65, 55, and 30. Mandell (101) stated that a *Formulation Test*, consisting of 15 items requiring a narrative statement to be translated into an algebraic equivalent, differentiated between research and nonresearch personnel.

An aptitude test for the selection of research personnel described by Weislogel (158) was based on a determination of critical requirements for successful participation in research and engineering work. Items were written to predict specific behaviors identified by scientists as crucial.

The summary by Stuit and others (143) provided, for each of several professional fields including engineering, law, medicine, dentistry, and nursing, a review of the research findings in the area as well as a statement of implications for counseling. Lannholm and Schrader (88) evaluated the effectiveness of the *Graduate Record Examinations*. Their review included not only reports of validity studies for graduate students in general, but also detailed statistical findings in many subjectmatter fields, e.g., chemistry, English, and history.

### General Aptitude Test Battery

Despite its importance, the *General Aptitude Test Battery* of the United States Employment Service has been infrequently mentioned in the literature during the past three years. The available information about this battery, and especially the published evidence concerning its demonstrated empirical validity for the predictive purposes for which it is used, remain distinctly inadequate.

The wide use of the GATB was indicated by the report of Petrullo, Cohen, and Meigh (126); in 1949 it was being administered in local offices of the U. S. Employment Service to 100,000 persons per year. This article and also that of Odell (119) described a research program being carried on thru cooperative relationships with various universities. Many of the projects were concerned with norms for special groups such as prepharmacy students; none was concerned with a follow-up validity study. One of the cooperative research projects was that of Taylor and others (147) at the University of Utah. The purpose was to expand upon the occupational aptitude pattern norms originally reported for the GATB. The end goal was one general college aptitude pattern plus academic area patterns for biology, chemistry, education, engineering, social science, medicine, and pharmacy. Samples studied in the different areas ranged in size from 49 in medicine to 123 in education. The "best" set of aptitudes for each of the seven areas and for general college all included G (intelligence) and V (verbal ability); N (numerical ability) was also represented for business, engineering, medicine, and pharmacy; S (spatial aptitude) for engineering and medicine; and Q (clerical perception) for education. Multiple correlations ranged from .41 to .63 with a median of .56. (These are not follow-up validity coefficients.) The overlapping of aptitudes for the areas reflected emphasis placed on establishing batteries that would identify all the academic areas in which a counselee could attain adequate success.

The Ohio State Employment Service testing staff (120) reported a study in which the GATB was administered to 439 high-school seniors in five northern Ohio schools. By a study of the obtained score distributions it

was concluded that the battery appeared applicable for use with this type of population. Perhaps most significant in the report was the statement that further research was needed to determine how well the test results have aided in the vocational adjustment of these high-school youth.

### Differential Prediction or Classification

The work of the past three years in the area of differential prediction or classification was keynoted by Thorndike (148). He stated that in its pure form, the problem is to determine which job is to be filled by which individual when all job applicants are to be divided among a given number of job categories. Thorndike went on to discuss the design, choice, and weighting of tests in a differential battery and pointed out the desirability of using simple, factorially pure tests, since these may be expected to have a wide range of validities for different job categories. French's monograph (53) may appropriately be mentioned here, since it provided a summary of data on the factorial composition of test scores, for studies in which rotations of axes were made.

Wesman and Bennett (161) stated three pertinent statistical principles: (a) If a test correlates to about the same extent with two criteria, it will be ineffective for direct prediction of differences; (b) If criteria are highly intercorrelated, small opportunity exists for differential prediction; and (c) Any difference is less reliable than the original measures upon which it is based. Mollenkopf (110) analyzed the problem of differential prediction existing when K tests are given to N individuals for each of whom there are criterion measures in two fields. The differential validity of the battery was shown to be a function of the multiple correlations of the battery with each criterion, the criterion intercorrelation, and the correlation between predicted scores. Mollenkopf (112) further considered problems in differential prediction, stressing particularly the critical importance of the magnitude of the predicted-score intercorrelation. Numerical examples were presented to illustrate the properties required in a test for it to be effective differentially. Brogden (20) demonstrated that a battery of tests with differential weighting for each job would yield a material increase in efficiency of selection over that afforded by a single predictor when people were hired from the same population of applicants for a number of jobs.

Several significant studies of the *Differential Aptitude Tests* have been reported. In one of these Doppelt and Bennett (42) examined the consistency of measurement by this battery for a group of students tested in Grade IX and retested in Grade XII. Correlations between corresponding scores ranged from .62 to .85, the highest being for verbal reasoning. That differences between test scores also were fairly consistent was demonstrated by correlating the difference between scores on two tests in 1947 with the corresponding difference in 1950. The median for 28 correlations of such differences was .50, N being 323.

Doppelt and Wesman (44) reported results of two validity studies of the DAT. In the first of these, six scores on the DAT given in November

1948, were correlated with 10 scores on the *Iowa Tests of General Educational Development* given in September 1949, grade by grade, with N's ranging from 44 to 66. For five out of six groups, DAT Numerical Ability correlated .80 or higher with TGED Quantitative Thinking; correlations of DAT Sentences with Correctness and Appropriateness of Expression ranged from .57 to .89; those between Verbal Reasoning and General Vocabulary ranged from .69 to .88. Some coefficients were surprising: the DAT Numerical Ability score correlated .71 with the TGED Correctness and Appropriateness of Expression score. The authors' second study involved 106 boys and 136 girls who were given the DAT in 1947 while in Grade IX and the *Essential High School Content Battery* in 1950. Over the three-year period, the DAT Verbal Reasoning and Sentences Tests accounted for 8 out of the 10 highest coefficients with the achievement measures, there being, for example, a correlation of .75 between Verbal Reasoning and EHSCB total score for the boys.

A follow-up study in six communities of 2900 students who had taken the DAT in 1947 was described by Bennett, Seashore, and Wesman (11) and Wesman (160). The 1700 usable replies to questionnaires were sorted according to post high-school career and percentile equivalents of average scores on the various tests obtained for these groups. Availability of test scores made by persons continuing in various fields will enable a comparison of a student's scores with those, say, for premedical students or general office clerks when these men were in high school. Three further extensive research reports were issued for the DAT by the Psychological Corporation (13, 14, 15). Bennett, Seashore, and Wesman (12) provided a casebook for use with the DAT which was designed to aid counselors in schools to use the test profiles more effectively. Other studies involving the DAT were those of Fruchter (56), Townsend (149), and Williams (164).

### Prediction of Other Scholastic Achievement

Garrett (58) summarized studies reported between 1929 and 1944 of special aptitude tests as predictors of college achievement. Olander, Van Wagenen, and Bishop (121) constructed scales of quantitative information and of perception of quantitative relations for use with first-graders. In a follow-up study of 289 students, correlations of the order of .50 were observed with the *Unit Scales of Attainment* in problem-solving and fundamental operations.

Use of an index of industriousness to improve prediction of achievement in college courses in English was demonstrated by Krathwohl (84). When a group of 308 sophomores at the Illinois Institute of Technology was divided into "industrious," "normal," and "indolent" groups on the basis of indexes of industriousness, the predictions of achievement made for each group separately were better than those for the entire group.

The *Iowa Foreign Language Aptitude Test* yielded correlations from .39 to .56, with a median of .45, for six different freshman language courses at the University of Michigan, according to Wallace (157).

### Music and Art Tests

By giving a "tonette test" consisting of sight reading after eight periods of instruction, Manor (105) was able to secure a correlation of .41 with later instrumental achievement. Lehman (89) gave the *Kwalwasser-Dykema Music Tests* to 50 students on entrance at the Brockport (N. Y.) State Teachers College, and also gave the *Kwalwasser-Ruch Test of Musical Accomplishment* before and after a 16 weeks' music theory course. The K-D scores correlated only .02 with the difference between the two K-R scores.

In the ninth grade of a Toronto high school in which art is taken by all students, Barrett (8) found that girls scored significantly higher than boys on both the *McAdory Art Test* and the *Meier Art Judgment Test*. However, in a study by Prothro and Perry (132) of the revised *Meier Art Judgment Test*, no sex difference was observed when performances of 223 male high-school and college students in Louisiana were compared with those of 187 females. Anderson (3) pointed out that wide discrepancies sometimes occur between scores on the present forms of the Meier and McAdory tests given to the same individuals. Correlations between scores on the two tests were only .23 for 111 women and .24 for 65 men.

Whistler and Thorpe (163) provided a new *Musical Aptitude Test* intended for use in Grades IV thru X. It involves rhythm, pitch, and melody recognition and pitch discrimination.

### Clerical Tests

An excellent summary of validity studies of clerical tests was that of Carruthers (26). Information was provided as to group tested, the test used, the bibliographic reference, the criterion, the size of the group, and the observed validity. In a factor study of the scores of 194 high-school students who were given 17 clerical aptitude tests, Bair (4) found that the *Minnesota Clerical Test* was related positively to more general types of clerical aptitude tests than others in the battery.

Construction of a new test designed to measure the aptitude for writing clear and tactful business letters was described by Kriedt (85). A key was developed by analysis of responses of two groups of 100 insurance company correspondence clerks, with cross-validation. In a new group correlations were .38 with supervisory ratings, .30 with job level, and .41 with ratings and level combined.

Blakemore (18) reported a correlation of .62 between scores on the *Hay Number Perception Test* and the key strokes per minute in typing from rough to finished copy, for a group of 35 typists in a large New York bank. Corresponding correlations for the *Minnesota Clerical Test* were .62 for the Number Section and .54 for the Names Section. Miller (108) obtained correlations of .83 for 99 men and .85 for 91 women between scores on the *Hay Number Perception Test* and the *Minnesota Clerical Test*.

### Mechanical Ability Tests

Poruben (130, 131) described the validation of the AGO *Mechanical Aptitudes Test* for a group of 72 students in five curriculums in a Yonkers, N. Y., trade school. Various of the four parts of the test yielded correlations ranging from .42 to .54 with a composite of grades in technical subjects taken during Grades X and XI. A one-year follow-up of 105 freshmen at Ohio State University who took Form CC of the *Owens-Bennett Mechanical Comprehension Test* was described by Halliday, Fletcher, and Cohen (73). Correlation with first-quarter average grade was .42; for 79 students, the correlation with first-year grades was .40.

The problems connected with the use of apparatus tests—cost, maintenance, etc.—are well known. The success of Nesburg and Smith (118) in producing a paper-and-pencil test duplicating the psychomotor performance involved in the *Vector Complex Reactometer* is therefore noteworthy. Correlations between scores on the new test and on the *Reactometer* ranged from .69 to .84 for various test sequences and groups.

Owens (123) evaluated a new test of mechanical comprehension which was a Bennett-type test but more schematic and difficult than the *Bennett Form BB*, and composed of five- rather than three-choice items. For 107 engineering seniors the correlation with grades in theoretical and applied mechanics was .49 (corrected for restriction in range), and .41 with median grades in seven relevant courses (also corrected). Other tests in the area include Crawford and Crawford's *Small Parts Dexterity Test* (33) and the *Stromberg Dexterity Test* (142).

### Other Aptitude Tests

Quite a number of short studies involving use of tests for selection of workers in the trades and services areas have appeared in the past three years. Two reviews appeared, both by Ghiselli and Brown. The first (61) surveyed the literature on the effectiveness of tests for the selection of auto mechanics. The second (60) covered relationships between aptitude-test scores and measures of trainability.

Maslow (107) reported that the U. S. Civil Service Commission had developed a written test for selection of skilled and semiskilled workers in the Government Printing Office and Bureau of Engraving and Printing. Laney (87) found correlations of .49 for the *Bennett Mechanical Comprehension Test* and .40 for the *Minnesota Paper Form Board* with supervisors' ratings of 60 experienced appliance service workers. Littleton (95) found the validity of the *Bennett Test of Mechanical Comprehension* for predicting instructors' ratings in auto trade courses to be slightly higher than that for either the *SRA Mechanical Aptitudes* battery or the *California Prognostic Test of Mechanical Ability*. Martin (106) reported scores on tests of mechanical information and spatial relations correlated .31 to .58 with ratings of performance of 45 auto-mechanics students.

Ghiselli and Brown (59) in a study of 67 new taxicab drivers found

that scores on dotting and tapping tests correlated .35 and .47 with accidents during first five weeks of employment. The *Bennett Test of Mechanical Comprehension* was found to differentiate significantly groups of firemen ranked "high" and "low" by their captains, in a study by Wolff and North (165). Du Bois and Watson (45) constructed a special *Police Aptitude Test* for use in St. Louis, but neither it nor other measures used in the Police Academy gave significant correlations with later on-the-job ratings.

The effectiveness of test data for vocational and educational guidance purposes is one of the most challenging problems in the field of testing. Barnette (6, 7) followed up cases of veterans who had completed the VA-sponsored advisement process at the New York City YMCA Vocational Service Center; the 890 replies received from some 1375 questionnaires sent out over a year after the last case was counseled were sorted by occupational field and into "success" and "failure" groups, "success" involving actually beginning the appropriate job, being satisfied with it, and continuing with it. Test scores for these groups were then compared for those in engineering work, salesmen, accountants, and clerical workers, and significant differences noted.

Despite changes in the applicant population and in the reasons for elimination, the Air Force pilot stanine was reported by Levine and Tupes (94) to have continued to be effective for predicting elimination from pilot training, the biserial between stanine and graduation-elimination being .57 for all reasons of elimination and .60 for flying deficiency alone.

New tests in the area included the *Aptitude Tests for Occupations*, by Roeder and Graham (135); *SET-Short Employment Tests*, by Bennett and Gelink (10); the *Store Personnel Test*, Form FS, by Seashore and Orbach (138); the *Aptitudes Associates Test of Sales Aptitude*, by Bruce (24); and the *Test for Ability To Sell* (Form 2), by Moss (117).

### Test Validity and the Criterion

A criticism of the tendency to build new tests without adequately taking into account what is already known about prediction of academic success was voiced by Travers (150). He maintained that it might be more profitable to devote time to a study of the criterion than to the proliferation of new tests which are somehow hoped will be more valid than previous ones. Travers and Wallace (153) pointed out that inconsistencies in validity may arise from the process of selection.

Fiske (51) discussed the question of selection of criteria and stressed the role of value judgments in the establishment of objectives. Wallace and Twichell (156) pointed out that the validity of a test used in industry might be affected by administrative procedures of the company. Adkins (1) stressed the value of objective performance measures and discussed the use of observational technics as measures of what one is trying to predict. The "dollar criterion," an over-all measure of worker effectiveness, involving converting production units, errors, time consumed, and similar factors into dollar units, was presented by Brogden and Taylor (22).

The considerable danger that may be involved in substitution of one criterion for another was pointed out by Severin (139). A related point was made by Anastasi (2), who stressed that validity was not simply a function of the test but of the use to which it was put.

One of the most important contributions in this area was Gulliksen's penetrating discussion of intrinsic validity (70). He pointed out that while in the early stages of a science it was appropriate for the scientist to be sure his measurements were at least as accurate as the results of skilled but nonscientific appraisal, at some point in the advance of psychology as a science it would seem appropriate for the psychologist to lead the way in establishing good criterion measures. Gulliksen also pointed out, apropos of coaching for predictive tests, that if there is a direct and causal relationship between an aptitude test and a criterion, it is likely that efforts to improve one's test score will also improve criterion performance; but if the test has only an indirect and not an intrinsic validity, then coaching will destroy the validity.

Criterion analysis thru application of the hypothetico-deductive method to factor analysis was advocated by Eysenck (48). Lubin (98) presented an outline of the algebraic procedure involved in Eysenck's method.

A demonstration of the pitfalls involved in using item-analysis data for a group, keying the items on this basis, and then estimating the validity for the same group was provided by Cureton (38), whose interpretation of such a coefficient was, "Baloney!" Further discussion of the need for and means of cross-validation was given in a series of papers by Mosier (116), Cureton (37), Katzell (83), and Wherry (162). Baker (5) recommended the use of compound rather than joint probability in the selection of items in the case of double cross-validation studies.

### Methods of Test Selection

During the past three years a number of new methods have been proposed for coping with the problem of selecting tests to form the most effective predictive battery.

Horst (78) provided a method for determining what validity an experimental test must have to make a specified increase in the predictive efficiency of a given test battery and developed (77) a solution for the problem of how long each test in a battery should be so that the correlation of the battery with the criterion will be a maximum. Taylor (146) also provided a solution for the allotment of time to the various tests, mathematically equivalent to that of Horst.

The selective efficiency of a test battery was expressed by Sichel (140) in terms of the "applicant's operating characteristic" and the "selector's operating characteristic." Summerfield and Lubin (144) presented a new procedure for selecting the minimum number of effective independent variables in a multiple-regression problem. The authors stated that their method provided a better decision procedure for ending the process of selection of tests than that of Wherry.

A coefficient of selection efficiency useful when applied to problems involving the validity of dichotomous predictors, or continuous predictors at various points of cut, was derived by Brogden (21). Brokaw (23) tested the hypothesis that predictive tests of high reliability and substantial validity might, when used in a battery, be considerably shortened without serious damage to battery validity.

### Item-Selection Procedures

A comprehensive review of the suggestions made over the past 50 years with regard to use of quantitative data on difficulty and discriminating power of test items was provided by Davis (40).

Defining the ability underlying a test as the common factor of item tetrachoric correlations corrected for guessing, Lord (96) derived an expression for the curvilinear relation between test score and this ability. It was indicated that reliability and this curvilinear correlation will be maximized by (a) minimizing variability of item difficulty; and (b) making the level of item difficulty somewhat easier than the halfway point between a chance percentage of correct answers and 100 percent correct. Similar conclusions were reached by Cronbach and Warrington (35) when they indicated that for item intercorrelations of the magnitude ordinarily encountered, narrowing the range of item difficulties will generally have beneficial effects on the validity of tests, and that a test designed to reject the lowest  $F$  percent should have items on the average at or above the threshold for men whose true ability is at the  $F$ th percentile.

Two solutions were presented by Bedell (9) to the problem of which items to discard, on the basis of item analysis, when revising a test designed to measure a single ability. French (54) derived a formula for keying a multiple-choice test for which no *a priori* key exists. Gleser and Du Bois (66) provided what they considered was a practical means of selecting items for a test so that it would yield the maximum correlation with the criterion. Levine (93) described a procedure whereby one might hope to be successful in the quest for that will-o'-the-wisp, the suppressor test.

A study by Ebel (46) of the reliability of item-discrimination data for a vocabulary test and for a test of basic skills in mathematics indicated that for these tests samples of 100 papers could be expected to provide indices of discrimination having a reliability over .80. Kuang (86) compared three item-analysis technics—biserials, Davis'  $z$ -transformations, and probit analysis—using a sample of 134 graduate students at Minnesota who took a 75-item test in statistics. When "best" sets of 10, 20, 30, and 40 items were selected by each method, agreement rose from 40 percent common items by all methods for 10-item tests to 75 percent for 40-item tests. Davis' method took least time, and probit analysis the most.

A similar study was that of Ely (47), who used four methods: that of Davis, Lawshe's  $D$ -values, phi coefficients, and percent high minus percent low passing the item. Six different-sized pairs of item-analysis groups

ranging from 10 percent to 50 percent of a total of 500 Purdue students were used to select from a pool of 150 vocabulary items four tests ranging in length from 20 to 80 items. While there was a statistically significant difference between the reliability in a new group of 183 students of tests derived by using the percent method from those by other methods, Jurgensen (82) pointed out that the difference was so small as to be of little practical significance.

Gulliksen (69) derived item indices which should remain relatively invariant with respect to changes in group mean and standard deviation. Johnson (80) proposed a new index of item validity, the U-L Index. Herfindahl (75) recommended the use of chi-square as a simple tool for selecting items, easily computed and used by a teacher.

An equation for predicting the effect of chance success on item-test correlation and on test reliability was derived by Plumlee (128). Predicted values were compared with empirical values in an experiment which used "identical" test items in multiple-choice and in answer-only (completion) form. Mollenkopf (109) found that whereas changing item placement had but slight effect on item indices in a power situation, both difficulty indices and item-test correlations were seriously affected when drop-out was high.

Using the responses of students in three samples of 370 each, Doppelt and Potts (43) studied the constancy of item-test coefficients estimated from Flanagan's table for 150 general information items. The coefficients were found to have standard errors only slightly larger than those for biserials computed for the same samples.

### Reliability and Standard Error of Measurement

A critical discussion of and a psychological rationale for the concepts of reliability and homogeneity were provided by Coombs (31). Cronbach (34) showed that coefficient alpha, a special case of which is the Kuder-Richardson coefficient of equivalence, was the mean of all split-half coefficients from different possible splittings of a test.

In an empirical study of the effect upon obtained reliability coefficients of several methods of splitting tests and of sampling variations, Clark (28) found that the subjects who happened to be used were an important cause for instability of reliability coefficients, whereas the method of splitting the test, if longitudinal, was not important. In another empirical study, several methods of estimating test reliability—the split-half, Guttman's  $L_4$ , and Kuder-Richardson Cases III and IV—together with Loevinger's estimate of homogeneity were compared by Gage and Damrin (57). Slight and unimportant differences among methods were found. Reliability was observed to increase with number of choices, especially from two to four. However, it was also shown that addition of choices might lower reliability if the test thus became inappropriate in difficulty for the group tested.

Horst (76) provided a formula for estimating total test reliability when scores were available for two parts comparable in all respects save length. Gulliksen (71) presented several methods for estimating the reliability

of a partially speeded test without the use of a parallel form. Cronbach and Warrington (36) further discussed the problem of estimating the reliability of speeded tests and provided an index of the degree of speeding. Essentially, a test was considered unspeeded when no subject's relative standing would be altered if he were given additional time on the test.

An equation was derived by Mollenkopf (113) for predicting the standard error of measurement at various points in the test-score distribution from the first four moments of the distribution and the matched-halves reliability. Green (67) proposed a criterion for determining the significance of the differences between the standard errors of measurement observed when a test has been given to more than one group of individuals. Woodbury (166) defined a new descriptive parameter of a test, its standard length, an invariant quantity as length is increased. A test with a reliability of .5 has a length equal to the standard length.

### Scoring

In three articles the problem of correction for chance success was considered. Hamilton (74) maintained that the usual correction-for-chance formula  $S=R - W/k - 1$  was improper, and he presented a formula for estimating real scores on a multiple-choice test from the raw scores. However, Lyerly (99) demonstrated that the usual formula yielded a close approximation to the maximum-likelihood estimate of an individual's true score on a test, and in criticism of Hamilton's method, indicated one of its consequences to be that the subject's estimated score would depend upon the distribution of scores in the group in which he happens to be tested. On the basis of an empirical study of item-analysis data for six pretests of varying levels of difficulty, Bryan, Burke, and Stewart (25) recommended that correction for guessing be employed in the scoring of pretests.

### Factors Related to Test Scores

A number of studies have appeared which involve the common element of some factor or factors related to test performance. For example, Doppelt (41) observed that psychology majors found both "science" and "non-science" items in Form G of the *Miller Analogies Test* easier than did individuals with other majors, and that science majors excelled nonscience majors on both types in terms of average item difficulties. However, the average of the item-test correlations did not differ much from group to group.

The question of whether speeding a test makes the scores reflect something different from what the scores would indicate when subjects are given plenty of time was studied by Mollenkopf (111). For a verbal antonyms test the rankings of students under the two conditions were practically the same. However, added time did tend to change the rankings for a mathematical aptitude test.

Davenport (39) related mean test scores by states on the *Army-Navy*

*Qualifying Examination* to variables reflecting "goodness of living" within the state. High relationships were observed between the state means and auto registrations, residents per 100,000 in *Who's Who*, and telephones per 1000 residents. Fruchter (55) pointed out that wrongs or error scores on tests, such as error in plotting accuracy and scale reading, were measures of carefulness.

The need for sufficient fore exercises to insure adequate comprehension of the analogy type of problem was stressed by Levine (90), who also proposed (91) a correction of special ability test scores for general ability. Schultz (136) examined performances on three mathematics tests of the *College Entrance Examination Board* in terms of amount and recency of training, and found these positively related to scores on the mathematics part of the *Scholastic Aptitude Test*.

After classifying mathematics items in each of three tests as "verbal" or "nonverbal" in terms of their manner of presentation, Plumlee (129) obtained correlations of each of these categories with scores on a verbal aptitude test. The correlations were not consistently different.

### General Procedures of Test Development

Two articles were concerned with general aspects of test construction. Flanagan (52) maintained that during the past 25 years, most test development work has been at the level of the technician, and urged that instead there be a more rational approach with emphasis on clear and precise definitions of what is to be measured and explicit hypotheses (termed rationales) regarding the behavior to be predicted. Test items then would be prepared to fit these rationales. A similar point of view was expressed by Travers (151), who contrasted the technician's approach with what he termed the "rational hypothesis" approach. In the latter, only items which were rationally hypothesized as belonging would be included in a scale.

### General Aspects of Mental Test Theory

A number of distinct contributions have been made during the three-year period in the field of test theory. Most outstanding of these was Gulliksen's *Theory of Mental Tests* (72). Coombs (32) developed a new scale for use in psychological work which does not involve a unit of measurement. This scale, which he termed an "ordered metric," falls logically between an interval and an ordinal scale. In two articles Comrey (29, 30) discussed logic and nature of measurement with regard to mental testing. In a series of articles (62, 63, 65) Glaser presented the concepts of multiple-operation measurement as applied to psychological tests. A subject's test score is defined as the mean of "inconsistent responses" on two or more administrations of a test, items in which are spaced along a scale such that the subject passes one or more items at one end and fails one or more at the other end.

## Bibliography

1. ADKINS, DOROTHY C. "Principles Underlying Observational Techniques of Evaluation." *Educational and Psychological Measurement* 11: 29-51; Spring 1951.
2. ANASTASI, ANNE. "The Concept of Validity in the Interpretation of Test Scores." *Educational and Psychological Measurement* 10: 67-78; Spring 1950.
3. ANDERSON, ROSE G. "A Note on the McAdory and Meier Art Tests in Counseling." *Educational and Psychological Measurement* 11: 81-86; Spring 1951.
4. BAIR, JOHN T. "Factor Analysis of Clerical Aptitude Tests." *Journal of Applied Psychology* 35: 245-49; August 1951.
5. BAKER, PAUL C. "Combining Tests of Significance in Cross-Validation." *Educational and Psychological Measurement* 12: 300-306; Summer 1952.
6. BARNETTE, W. LESLIE, JR. "Occupational Aptitude Pattern Research." *Occupational* 29: 5-12; October 1950.
7. BARNETTE, W. LESLIE, JR. "Occupational Aptitude Pattern of Selected Groups of Counseled Veterans." *Psychological Monographs*. Washington, D. C.: American Psychological Association, 1951. 49 p.
8. BARRETT, HARRY O. "Sex Differences in Art Ability." *Journal of Educational Research* 43: 391-93; January 1950.
9. BEDELL, B. J. "Determination of the Optimum Number of Items To Retain in a Test Measuring a Single Ability." *Psychometrika* 15: 419-30; December 1950.
10. BENNETT, GEORGE K., and GELINK, MARJORIE. *The Short Employment Tests*. New York: Psychological Corp., 1951.
11. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. "Aptitude Testing: Does It 'Prove Out' in Counseling Practice?" *Occupations* 30: 584-93; May 1952.
12. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *Counseling from Profiles: A Casebook for the Differential Aptitude Tests*. New York: Psychological Corp., 1951. 95 p.
13. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *The Differential Aptitude Tests*. Fifth Research Report. New York: Psychological Corp., 1951. 24 p.
14. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *Validation of the Differential Aptitude Tests*. Fourth Research Report. New York: Psychological Corp., 1951. 8 p.
15. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *Validation of the Differential Aptitude Tests*. Third Research Report. New York: Psychological Corp., 1949. 37 p.
16. BIRDIE, RALPH F. "The Differential Aptitude Tests as Predictors in Engineering Training." *Journal of Educational Psychology* 42: 114-23; February 1951.
17. BIRDIE, RALPH F., and SUTTER, NANCY A. "Predicting Success of Engineering Students." *Journal of Educational Psychology* 41: 184-90; March 1950.
18. BLAKEMORE, ARLINE. "Reducing Typing Costs with Aptitude Tests." *Personnel Journal* 30: 20-24; May 1951.
19. BRADEN, GEORGE D. "Use of the Law School Admission Test at the Yale Law School." *Journal of Legal Education* 3: 202-206; Winter 1950.
20. BROGDEN, HUBERT E. "Increased Efficiency of Selection Resulting from Replacement of a Single Predictor with Several Different Predictors." *Educational and Psychological Measurement* 11: 173-95; Summer 1951.
21. BROGDEN, HUBERT E. "A New Coefficient: Application to Biserial Correlation and to Estimation of Selective Efficiency." *Psychometrika* 14: 169-82; September 1949.
22. BROGDEN, HUBERT E., and TAYLOR, ERWIN K. "The Dollar Criterion—Applying the Cost Accounting Concept to Criterion Construction." *Personnel Psychology* 3: 133-54; Summer 1950.
23. BROKAW, LELAND D. "Comparative Validities of 'Short' Versus 'Long' Tests." *Journal of Applied Psychology* 35: 325-30; October 1951.
24. BRUCE, MARTIN M. *Aptitudes Associates Test of Sales Aptitude*. New York: the Author (524 East 20th Street), 1950.
25. BRYAN, MIRIAM M.; BURKE, PAUL J.; and STEWART, NAOMI. "Correction for Guessing in the Scoring of Pretests: Effect upon Item Difficulty and Item Validity Indices." *Educational and Psychological Measurement* 12: 45-56; Spring 1952.

26. CARRUTHERS, JOHN B. "Tabular Summary Showing Relation Between Clerical Test Scores and Occupational Performance." *Occupations* 29: 40-50; October 1950.
27. CHAUNCEY, HENRY. "The Use of the Selective Service College Qualification Test in the Deferment of College Students." *Science* 116: 73-79; July 25, 1952.
28. CLARK, EDWARD L. "Methods of Splitting vs. Samples as Sources of Instability in Test-Reliability Coefficients." *Harvard Educational Review* 19: 178-82; May 1949.
29. COMREY, ANDREW L. "Mental Testing and the Logic of Measurement." *Educational and Psychological Measurement* 11: 323-33; Autumn 1951.
30. COMREY, ANDREW L. "An Operational Approach to Some Problems in Psychological Measurement." *Psychological Review* 57: 217-28; July 1950.
31. COOMBS, CLYDE H. "The Concepts of Reliability and Homogeneity." *Educational and Psychological Measurement* 10: 43-56; Spring 1950.
32. COOMBS, CLYDE H. "Psychological Scaling Without a Unit of Measurement." *Psychological Review* 57: 145-58; May 1950.
33. CRAWFORD, JOHN E., and CRAWFORD, DOROTHEA M. *Small Parts Dexterity Test*. New York: Psychological Corp., 1949.
34. CRONBACH, LEE J. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16: 297-334; September 1951.
35. CRONBACH, LEE J., and WARRINGTON, WILLARD G. "Efficiency of Multiple-Choice Tests as a Function of the Spread of Item Difficulties." *Psychometrika* 17: 127-47; June 1952.
36. CRONBACH, LEE J., and WARRINGTON, WILLARD G. "Time-Limit Tests: Estimating Their Reliability and Degree of Speeding." *Psychometrika* 16: 167-88; June 1951.
37. CURETON, EDWARD E. "The Need and Means of Cross-Validation. II. Approximate Linear Restraints and Best Predictor Weights." *Educational and Psychological Measurement* 11: 12-15; Spring 1951.
38. CURETON, EDWARD E. "Validity, Reliability, and Baloney." *Educational and Psychological Measurement* 10: 94-96; Spring 1950.
39. DAVENPORT, KENNETH S., and REMMERS, HERMANN H. "Factors in State Characteristics Related to Average A-12 V-12 Test Scores." *Journal of Educational Psychology* 41: 110-15; February 1950.
40. DAVIS, FREDERICK B. "Item Analysis in Relation to Educational and Psychological Testing." *Psychological Bulletin* 49: 97-121; March 1952.
41. DOPPELT, JEROME E. "Difficulty and Validity of Analogies Item in Relation to Major Field of Study." *Journal of Applied Psychology* 35: 30-33; February 1951.
42. DOPPELT, JEROME E., and BENNETT, GEORGE K. "A Longitudinal Study of the Differential Aptitude Tests." *Educational and Psychological Measurement* 11: 228-37; Summer 1951.
43. DOPPELT, JEROME E., and POTTS, EDITH M. "The Constancy of Item-Test Correlation Coefficients Computed from Upper and Lower Groups." *Journal of Educational Psychology* 40: 378-81; October 1949.
44. DOPPELT, JEROME E., and WESMAN, ALEXANDER G. "The Differential Aptitude Tests as Predictors of Achievement Test Scores." *Journal of Educational Psychology* 43: 210-17; April 1952.
45. DU BOIS, PHILIP H., and WATSON, ROBERT I. "The Selection of Patrolmen." *Journal of Applied Psychology* 34: 90-95; April 1950.
46. EBEL, ROBERT L. "The Reliability of an Index of Item Discrimination." *Educational and Psychological Measurement* 11: 403-408; Autumn 1951.
47. ELY, JEROME H. "Studies in Item Analysis 2: Effects of Various Methods upon Test Reliability." *Journal of Applied Psychology* 35: 194-203; June 1951.
48. EYSENCK, HANS J. "Criterion Analysis—An Application of the Hypothetico-Deductive Method to Factor Analysis." *Psychological Review* 57: 38-53; January 1950.
49. FEENEY, BERNARD J. "How Good Are Legal Aptitude Tests?" *Journal of Legal Education* 4: 69-85; Autumn 1951.
50. FINDLEY, WARREN G. "The Selective Service College Qualification Test." *American Psychologist* 6: 181-83; May 1951.
51. FISKE, DONALD W. "Values, Theory, and the Criterion Problem." *Personnel Psychology* 40: 93-98; Spring 1951.

52. FLANAGAN, JOHN C. "The Use of Comprehensive Rationales in Test Development." *Educational and Psychological Measurement* 11: 151-55; Spring 1951.
53. FRENCH, JOHN W. *Description of Aptitude and Achievement Tests in Terms of Rotated Factors*. Psychometric Monograph No. 5. Chicago: University of Chicago Press, 1951. 278 p.
54. FRENCH, JOHN W. "A Technique for Criterion-Keying and Selecting Test Items." *Psychometrika* 17: 101-106; March 1952.
55. FRUCHTER, BENJAMIN. "Error Scores as a Measure of Carefulness." *Journal of Educational Psychology* 41: 279-91; May 1950.
56. FRUCHTER, BENJAMIN. "Orthogonal and Oblique Solutions of a Battery of Aptitude, Achievement and Background Variables." *Educational and Psychological Measurement* 12: 20-38; Spring 1952.
57. GAGE, NATHANIEL L., and DAMRIN, DORA E. "Reliability, Homogeneity and Number of Choices." *Journal of Educational Psychology* 41: 385-404; November 1950.
58. GARRETT, HARLEY F. "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Science and Teachers Colleges." *Journal of Experimental Education* 18: 91-138; December 1949.
59. GHISELLI, EDWIN E., and BROWN, CLARENCE W. "The Prediction of Accidents of Taxicab Drivers." *Journal of Applied Psychology* 33: 540-46; December 1949.
60. GHISELLI, EDWIN E., and BROWN, CLARENCE W. "Validity of Aptitude Tests for Predicting Trainability of Workers." *Personnel Psychology* 4: 243-60; Autumn 1951.
61. GHISELLI, EDWIN E., and BROWN, CLARENCE W. "Validity of Tests for Auto Mechanics." *Journal of Applied Psychology* 35: 23-24; February 1951.
62. GLASER, ROBERT. "The Application of the Concepts of Multiple-Operation Measurement to the Response Patterns on Psychological Tests." *Educational and Psychological Measurement* 11: 372-82; Autumn 1951.
63. GLASER, ROBERT. "Multiple Operation Measurement." *Psychological Review* 42: 241-53; July 1950.
64. GLASER, ROBERT. "Predicting Achievement in Medical School." *Journal of Applied Psychology* 35: 272-74; August 1951.
65. GLASER, ROBERT. "The Reliability of Inconsistency." *Educational and Psychological Measurement* 12: 60-64; Spring 1952.
66. GLEESER, GOLDINE C., and DU BOIS, PHILIP H. "A Successive Approximation Method of Maximizing Test Validity." *Psychometrika* 16: 129-39; March 1951.
67. GREEN, BERT F., JR. "A Test of the Equality of Standard Errors of Measurement." *Psychometrika* 15: 251-57; September 1950.
68. GREGG, GEORGE W. "Investigation of the Reliability and Validity of the Engineering and Physical Science Aptitude Test." *Journal of Educational Research* 45: 299-305; December 1951.
69. GULLIKSEN, HAROLD. "Effect of Group Heterogeneity on Item Parameters." *Psychometrika* 16: 285-96; September 1951.
70. GULLIKSEN, HAROLD. "Intrinsic Validity." *American Psychologist* 5: 511-17; October 1950.
71. GULLIKSEN, HAROLD. "The Reliability of Speeded Tests." *Psychometrika* 15: 259-69; September 1950.
72. GULLIKSEN, HAROLD. *Theory of Mental Tests*. New York: John Wiley and Sons, 1950, 486 p.
73. HALLIDAY, ROBERT W.; FLETCHER, FRANK M., JR.; and COHEN, RITA M. "Validity of the Owens-Bennett Mechanical Comprehension Test." *Journal of Applied Psychology* 35: 321-24; October 1951.
74. HAMILTON, C. HORACE. "Bias and Error in Multiple-Choice Tests." *Psychometrika* 15: 151-68; June 1950.
75. HERFINDAHL, ORRIS C. "An Application of Chi-Square to the Determination of the Discriminating Power of Test Questions." *Journal of Educational Psychology* 40: 371-77; October 1949.
76. HORST, PAUL. "Estimating Total Test Reliability from Parts of Unequal Length." *Educational and Psychological Measurement* 11: 368-71; Autumn 1951.
77. HORST, PAUL. "Optimal Test Length for Maximum Battery Validity." *Psychometrika* 16: 189-202; June 1951.
78. HORST, PAUL. "The Relationship Between the Validity of a Single Test and Its Contribution to the Predictive Efficiency of a Test Battery." *Psychometrika* 16: 57-66; March 1951.

79. JOHNSON, A. PEMBERTON. "The Development and Use of Law Aptitude Tests." *Journal of Legal Education* 3: 192-201; Winter 1950.
80. JOHNSON, A. PEMBERTON. "Notes on a Suggested Index of Item Validity: The U-L Index." *Journal of Educational Psychology* 42: 499-504; November 1951.
81. JOHNSON, A. PEMBERTON. "Tests and Testing Programs of Interest to Engineering Education." *Journal of Engineering Education* 41: 277-83; January 1951.
82. JURGENSEN, CLIFFORD E. "A Note on Ely's 'Effects of Various Methods upon Test Reliability'." *Journal of Applied Psychology* 35: 204; June 1951.
83. KATZELL, RAYMOND A. "The Need and Means of Cross-Validation: III. Cross-Validation of Item Analyses." *Educational and Psychological Measurement* 11: 16-22; Spring 1951.
84. KRATHWOHL, WILLIAM C. "Relative Contributions of Vocabulary and an Index of Industriousness for English to Achievement in English." *Journal of Educational Psychology* 42: 97-104; February 1951.
85. KRIEDT, PHILIP H. "Validation of a Correspondence Aptitude Test." *Journal of Applied Psychology* 36: 5-7; February 1952.
86. KUANG, H. P. "A Critical Evaluation of the Relative Efficiency of Three Techniques in Item Analysis." *Educational and Psychological Measurement* 12: 248-66; Summer 1952.
87. LANEY, ARTHUR R., JR. "Validity of Employment Tests for Gas-Appliance Service Personnel." *Personnel Psychology* 4: 199-208; Summer 1951.
88. LANNHOLM, GERALD V., and SCHRADER, WILLIAM B. *Predicting Graduate School Success*. Princeton, N. J.: Educational Testing Service, 1951. 50 p.
89. LEHMAN, CHARLES F. "An Investigation of Musical Achievement and Relationship to Intelligence and Musical Talent." *Journal of Educational Research* 45: 623-29; April 1952.
90. LEVINE, ABRAHAM S. "Construction and Use of Verbal Analogy Items." *Journal of Applied Psychology* 34: 105-107; April 1950.
91. LEVINE, ABRAHAM S. "Correcting Special Ability Test Scores for General Ability." *Journal of Applied Psychology* 33: 566-68; December 1949.
92. LEVINE, ABRAHAM S. "Minnesota Psycho-Analogies Test." *Journal of Applied Psychology* 34: 300-305; October 1950.
93. LEVINE, ABRAHAM S. "A Technique for Developing Suppression Tests." *Educational and Psychological Measurement* 12: 313-15; Summer 1952.
94. LEVINE, ABRAHAM S., and TUPES, ERNEST C. "Postwar Research in Pilot Selection and Classification." *Journal of Applied Psychology* 36: 157-60; June 1952.
95. LITTLETON, ISAAC T. "Prediction in Auto Trade Courses." *Journal of Applied Psychology* 36: 15-19; February 1952.
96. LORD, FREDERIC M. "An Investigation of the Relation of the Reliability of Multiple-Choice Tests to the Distribution of Item Difficulties." *Psychometrika* 17: 181-94; June 1952.
97. LORD, FREDERIC M.; COWLES, JOHN T.; and CYNAMON, MANUEL. "The Pre-Engineering Inventory as a Predictor of Success in Engineering Colleges." *Journal of Applied Psychology* 34: 30-39; February 1950.
98. LUBIN, ARDIE. "A Note on 'Criterion Analysis.'" *Psychological Review* 57: 54-57; January 1950.
99. LYERLY, SAMUEL B. "A Note on Correcting for Chance Success in Objective Tests." *Psychometrika* 16: 21-30; March 1951.
100. MANDELL, MILTON M. "The Administrative Judgment Test." *Journal of Applied Psychology* 34: 145-47; June 1950.
101. MANDELL, MILTON M. "Measuring Originality in the Physical Sciences." *Educational and Psychological Measurement* 10: 380-85; Autumn 1950.
102. MANDELL, MILTON M. "Scientific Selection of Engineers." *Personnel* 26: 296-98; January 1950.
103. MANDELL, MILTON M. "Selecting Chemists for the Federal Government." *Personnel Psychology* 3: 53-56; Spring 1950.
104. MANDELL, MILTON M., and CHAD, SEYMOUR. "Tests for Selecting Engineers." *Public Personnel Review* 11: 217-22; October 1950.
105. MANOR, HAROLD C. "A Study in Prognosis: The Guidance Value of Selected Measures of Musical Aptitude, Intelligence, Persistence, and Achievement in Tonette and Adaption Classes for Prospective Instrumental Students." *Journal of Educational Psychology* 41: 31-50; January 1950.

106. MARTIN, GLENN C. "Test Batteries for Trainees in Auto Mechanics and Apparel Design." *Journal of Applied Psychology* 35: 20-22; February 1951.
107. MASLOW, ALBERT P. "Written Tests To Select and Place Unskilled and Semi-Skilled Workers." *Public Personnel Review* 11: 96-99; April 1950.
108. MILLER, RICHARD B. "Reducing the Time Required for Testing Clerical Applicants." *Personnel Journal* 28: 364-66; March 1950.
109. MOLLENKOPF, WILLIAM G. "An Experimental Study of the Effects on Item-Analysis Data of Changing Item Placement and Test Time Limit." *Psychometrika* 15: 291-315; September 1950.
110. MOLLENKOPF, WILLIAM G. "Predicted Differences and Differences Between Predictions." *Psychometrika* 15: 409-17; December 1950.
111. MOLLENKOPF, WILLIAM G. "Slow, But How Sure?" *College Board Review* 11: 147-51; June 1950.
112. MOLLENKOPF, WILLIAM G. "Some Aspects of the Problem of Differential Prediction." *Educational and Psychological Measurement* 12: 39-44; Spring 1952.
113. MOLLENKOPF, WILLIAM G. "Variation of the Standard Error of Measurement." *Psychometrika* 14: 189-229; September 1949.
114. MOORE, JOSEPH E. "A Decade of Attempts To Predict Scholastic Success in Engineering Schools." *Occupations* 28: 92-96; November 1949.
115. MORRIS, WOODROW W. "Validity of the Professional Aptitude Test in Medicine." *Medical Education* 26: 56-58; January 1951.
116. MOSIER, CHARLES I. "Symposium: The Need and Means of Cross-Validation: I. Problems and Designs of Cross-Validation." *Educational and Psychological Measurement* 11: 5-11; Spring 1951.
117. MOSS, FRED A. *Test for Ability to Sell (Form 2)*. Washington, D. C.: George Washington University, Center for Psychological Service, 1950.
118. NESBERG, LLOYD S., and SMITH, KARL U. "Measurement of a Complex Psychomotor Performance by Means of a Printed Test." *Journal of Applied Psychology* 34: 309-12; October 1950.
119. ODELL, CHARLES E. "Cooperative Research in Aptitude Test Development." *Educational and Psychological Measurement* 9: 396-400; Autumn 1949.
120. OHIO STATE EMPLOYMENT SERVICE TESTING STAFF. "A General Aptitude Test Battery Study with High School Seniors." *Educational and Psychological Measurement* 9: 281-89; Autumn 1949.
121. OLANDER, HERBERT T.; VAN WAGENEN, MARVIN J.; and BISHOP, HELEN M. "Predicting Arithmetic Achievement." *Journal of Educational Research* 43: 66-73; September 1949.
122. OWENS, WILLIAM A. "An Aptitude Test for Veterinary Medicine." *Journal of Applied Psychology* 34: 295-99; October 1950.
123. OWENS, WILLIAM A., JR. "A Difficult New Test of Mechanical Comprehension." *Journal of Applied Psychology* 34: 77-81; April 1950.
124. PEARSON, JOHN S., and STRATE, MARVIN W. "The Minnesota Psycho-Analogies Test in the Selection of Psychologists for the Public Service." *Journal of Applied Psychology* 35: 314-15; October 1951.
125. PETERSON, SHAILER A. "Dental Aptitude Testing Program Will Become Nationwide for the 1951 Entrants." *College and University* 26: 112-14; October 1950.
126. PETRULLO, LUIGI; COHEN, IRVING; and MEIGH, CHARLES. "The Employment Service Testing Program." *Employment Security Review* 16: 17-19, 23; March 1949.
127. PIERSON, GEORGE A., and JEX, FRANK B. "Using the Cooperative General Achievement Tests To Predict Success in Engineering." *Educational and Psychological Measurement* 11: 397-402; Autumn 1951.
128. PLUMLEE, LYNNETTE B. "The Effect of Difficulty and Chance Success on Item-Test Correlation and on Test Reliability." *Psychometrika* 17: 69-86; March 1952.
129. PLUMLEE, LYNNETTE B. "The Verbal Component in Mathematics Items." *Educational and Psychological Measurement* 9: 679-84; Winter 1949.
130. PORUBEN, ADAM, JR. "Validation and Standardization of the AGO General Mechanical Aptitudes Test." *Journal of Psychology* 29: 133-55; January 1950.
131. PORUBEN, ADAM, JR. "Validation and Standardization of the AGO General Mechanical Aptitudes Test for the Selection of Civilian Employees in War Department Installations." *Educational and Psychological Measurement* 10: 254-62; Summer 1950.

132. PROTHRO, E. TERRY, and PERRY, HAROLD T. "Group Differences in Performance on the Meier Art Test." *Journal of Applied Psychology* 34: 96-97; April 1950.
133. RALPH, RAY B., and TAYLOR, CALVIN W. "A Comparative Evaluation of the Professional Aptitude Test and the General Aptitude Test Battery." *Journal of the Association of American Medical Colleges* 25: 33-40; January 1950.
134. RALPH, RAY B., and TAYLOR, CALVIN W. "The Role of Tests in the Medical Selection Program." *Journal of Applied Psychology* 36: 107-11; April 1952.
135. ROEDER, WESLEY S., and GRAHAM, HERBERT B. *Aptitude Tests for Occupations*. Los Angeles: California Test Bureau, 1951.
136. SCHULTZ, DOUGLAS G. "The Comparability of Scores from Three Mathematics Tests of the College Entrance Examination Board." *Psychometrika* 15: 369-84; December 1950.
137. SCHULTZ, DOUGLAS G. "The Relationship Between Scores on the Science Test of the Medical College Admission Test and Amount of Training in Biology, Chemistry, and Physics." *Educational and Psychological Measurement* 11: 138-50; Spring 1951.
138. SEASHORE, HAROLD G., and ORBACH, CHARLES E. *Store Personnel Test, Form FS*. New York: Psychological Corp., 1951.
139. SEVERIN, DARYL. "The Predictability of Various Kinds of Criteria." *Personnel Psychology* 5: 93-104; Summer 1952.
140. SICHEL, HERBERT S. "The Selective Efficiency of a Test Battery." *Psychometrika* 17: 1-39; March 1952.
141. STALNAKER, JOHN M. "Medical College Admission Test." *Journal of the Association of American Medical Colleges* 25: 428-34; November 1950.
142. STROMBERG, ELEROY L. *Stromberg Dexterity Test*. New York: Psychological Corp., 1951.
143. STUIT, DEWEY B., and OTHERS. *Predicting Success in Professional Schools*. Washington, D. C.: American Council on Education, 1949, 187 p.
144. SUMMERFIELD, A., and LUBIN, ARDIE. "A Square Root Method of Selecting a Minimum Set of Variables in Multiple Regression: I. The Method." *Psychometrika* 16: 271-84; September 1951.
145. TAYLOR, CALVIN W. "Check Studies on the Predictive Value of the MCAT." *Journal of the Association of American Medical Colleges* 25: 269-71; July 1950.
146. TAYLOR, CALVIN W. "Maximizing Predictive Efficiency for a Fixed Total Testing Time." *Psychometrika* 15: 391-406; December 1950.
147. TAYLOR, CALVIN W., and OTHERS. "General Aptitude Test Battery Patterns for College Areas." *Occupations* 29: 518-26; April 1951.
148. THORNDIKE, ROBERT L. "The Problem of Classification of Personnel." *Psychometrika* 15: 215-35; September 1950.
149. TOWNSEND, AGATHA. "The Differential Aptitude Tests—Some Data on the Reliability and Intercorrelation of the Parts." *Educational Records Bulletin* 53: 39-47; January 1950.
150. TRAVERS, ROBERT M. W. "Prediction of Achievement." *School and Society* 70: 293-94; November 5, 1949.
151. TRAVERS, ROBERT M. W. "Rational Hypotheses in the Construction of Tests." *Educational and Psychological Measurement* 11: 128-37; Spring 1951.
152. TRAVERS, ROBERT M. W., and WALLACE, WIMBURN L. "The Assessment of the Academic Aptitude of the Graduate Student." *Educational and Psychological Measurement* 10: 371-79; Autumn 1950.
153. TRAVERS, ROBERT M. W., and WALLACE, WIMBURN L. "Inconsistency in the Predictive Value of a Battery of Tests." *Journal of Applied Psychology* 34: 237-39; August 1950.
154. TRAXLER, ARTHUR E. "Objective Testing in the Field of Accounting." *Educational and Psychological Measurement* 11: 427-39; Autumn 1951.
155. TREUMANN, MILDRED J., and SULLIVAN, BEN A. "Use of the Engineering and Physical Science Aptitude Test as a Predictor of Academic Achievement of Freshman Engineering Students." *Journal of Educational Research* 43: 129-33; October 1949.
156. WALLACE, S. RAINS, JR., and TWICHELL, CONSTANCE M. "Managerial Procedures and Test Validities." *Personnel Psychology* 2: 277-92; Autumn 1949.
157. WALLACE, WIMBURN L. "The Prediction of Grades in Specific College Courses." *Journal of Educational Research* 44: 587-97; April 1951.

158. WEISLOGEL, MARY H. *The Development of a Test for Selecting Research Personnel*. Pittsburgh: American Institute for Research, 1950. 33 p.
159. WEISS, IRVING. "Prediction of Academic Success in Dental School." *Journal of Applied Psychology* 36: 11-14; February 1952.
160. WESMAN, ALEXANDER G. "Guidance Testing." *Occupations* 30: 10-14; October 1951.
161. WESMAN, ALEXANDER G., and BENNETT, GEORGE K. "Problems of Differential Prediction." *Educational and Psychological Measurement* 11: 265-72; Summer 1951.
162. WHERRY, ROBERT J. "The Need for Cross-Validation: IV. Comparison of Cross-Validation with Statistical Inference of Betas and Multiple R from a Single Sample." *Educational and Psychological Measurement* 11: 23-28; Spring 1951.
163. WHISTLER, HARVEY S., and THORPE, LOUIS P. *Musical Aptitude Test*. Los Angeles: California Test Bureau, 1950.
164. WILLIAMS, NANCY. "A Study of the Validity of the Verbal Reasoning Subtest and the Abstract Reasoning Subtest of the Differential Aptitude Tests." *Educational and Psychological Measurement* 12: 129-31; Spring 1952.
165. WOLFF, W. M., and NORTH, ALVIN J. "Selection of Municipal Firemen." *Journal of Applied Psychology* 35: 25-29; February 1951.
166. WOODBURY, MAX A. "On the Standard Length of a Test." *Psychometrika* 16: 103-106; March 1951.

## CHAPTER IV

### Development and Applications of Nonprojective Tests of Personality and Interest

DAVID V. TIEDEMAN and KENNETH M. WILSON

THIS review concerns tests similar to those included in the "Character and Personality" and "Vocations-Interests" sections of Buros' *Third Mental Measurements Yearbook* (13). The *Wechsler-Bellevue Intelligence Scale* and multiple-choice versions of the Rorschach are excluded by this definition of nonprojective tests of personality and interests.

#### Trends and Developments

During the previous three-year period, Traxler and Jacobs (107) noted that the amount of research concerning older inventories like the *Bernreuter Personality Inventory*, the *Bell Adjustment Inventory*, and the *Allport-Vernon Study of Values* was less than that concerning newer inventories like the *Minnesota Multiphasic Personality Inventory* (MMPI). With the exception of the *Strong Vocational Interest Blank*, this trend continued into the current three-year period. The MMPI, the *Kuder Preference Record-Vocational*, and the *Strong Vocational Interest Blank* were the inventories studied most frequently during this period. Revisions of several older inventories, several new inventories, and several new scales for existing inventories appeared on the scene.

Except when inventories were rekeyed especially for the purpose, personality- and interest-inventory scores added little to the efficiency of aptitude and achievement measures for the prediction of educational success. It is interesting to note, however, that during this period many reliable differences in personality- and interest-inventory score patterns among various groups were found. This suggests that structured inventories may be more useful for inferring group membership than for inferring success within any one group.

Clinical and counseling psychologists continued their interest in the development of appropriate statistical models for their research problems, while the literature on Guttman's scale theory and Lazarsfeld's theory of latent structure continued to grow. Attention was given to multivariate analysis and its relation to profile interpretation.

#### Summaries

Abstracts of the employee selection work of 476 investigators, indexed by job title, author, and test, and describing subjects, criteria, validity, and reliability, were compiled by Dorcus and Jones (34).

Three issues of the *Annual Review of Psychology* were published during the current period, the first issue appearing in 1950. Altho different from

the *Review of Educational Research* in organization and emphasis, its content is somewhat similar, and many of the references included in the present review and its supplementary bibliography have been considered in the *Annual Review of Psychology* (3, 11, 23, 37, 41, 61, 69, 72, 99).

### Factor Studies of Personality and Interest

Some time ago, Cattell undertook the task of investigating the personality sphere thru factorial analyses of behavior rating, questionnaire, and objective test data. Results of analyses of the factorial content of questionnaire self-estimates (18) and of objective test data (17) published during this period, included the isolation of 19 oblique factors in the questionnaire data and 11 in the test data. Cattell and Saunders (22) attempted to match the factors from analyses of the three types of responses and isolated 12 factors. However, three rating factors, nine questionnaire factors, and three test factors were either unmatched or unrepresented.

In a paper on the ergic structure of man, Cattell (16) expressed dissatisfaction with present interpretation of basic human drives and initiated inquiry into this area within a framework of 23 hypothesized ergs and metanergs, terms that are functions of "drives" and "sentiments" respectively. He devised 50 attitude measures, at least two for each of the hypothesized variables, and analyzed them factorially. Seven definite ergs, the possibility of another erg, and one metanerg were indicated. These findings were integrated into a consistent framework in a book (19) that deserves attention.

Cattell's approach is refreshing and stimulating, not only because of the comprehensive nature of his investigations, but also because of the many new methods of personality assessment incorporated in his work (21).

Thurstone (102) reanalyzed the *Guilford Inventory of Factors*, STDCR, the *Guilford-Martin Inventory of Factors* GAMIN, and three additional scales, using reliability coefficients in the diagonal of the intercorrelation matrix, which, he emphasized, made his a first-order analysis, i.e., a verification study of tentatively established factors. The seven factors isolated were included in the *Thurstone Temperament Schedule* (103). In a second-order analysis (i.e., communalities in the diagonal) of these data, Baehr (1) found four second-order factors which were substantiated somewhat by an independent investigation using paired comparison ratings.

A factor analysis by Cottle (26) of the responses of 400 male veterans to the MMPI, the *Strong Vocational Interest Blank*, the *Kuder Preference Record-Vocational*, and the *Bell Adjustment Inventory* resulted in the isolation of seven interpretable factors, two largely from the personality inventories and five largely from the interest inventories. Little overlap of the personality and interest inventories was observed.

Wheeler, Little, and Lehner (112) and Tyler (108) studied the internal structure of the MMPI by factorial methods. In neither study were more than five factors isolated.

Vernon (109) selected 58 high-grade occupations and obtained for every pair an average of five judgments of similarity or dissimilarity on a seven-point scale. Analysis of the intercolumnar correlations of each pair of occupations resulted in the isolation of four bipolar factors: gregarious versus isolated, social welfare versus administrative, scientific versus display, and verbal versus active.

### New and Revised Inventories

During the period under consideration, the Guilford series of personality inventories was reduced to one form of 300 items and published as the *Guilford-Zimmerman Temperament Survey* (54). Areas surveyed, each by 30 items, are: general activity, restraint, ascendance, sociability, emotional stability, objectivity, friendliness, thoughtfulness, personal relations, and masculinity. The *Thurstone Temperament Schedule* (103), a 140-item test based on factor studies of Guilford's inventories, covering areas called active, vigorous, impulsive, dominant, stable, sociable, and reflective, was published.

The *S.R.A. Youth Inventory* by Remmers and Shimberg (83) and the *Heston Personal Adjustment Inventory* (58), both of which may be used with high-school pupils, were published, and the *Mooney Problem Check Lists*, Grades VII thru IX and X thru XII, were revised by Mooney and Gordon (75). Bell (2) published a 90-item *Personal Preference Inventory* yielding measures of maladjustment with respect to economic background, social attitudes, and masculinity-femininity.

Woodman (117), in an attempt at indirect measurement of students' attitudes toward academic success in college, developed "An Evaluation of Student Opinions" which, when combined with the *ACE Psychological Examination* and school grades, resulted in increased prediction of college achievement. A College Entrance Examination Board questionnaire designed by Myers and Schultz (78) to tap motivation for attending college, intellectual interests, teacher relations, and study habits added only slightly to the predictive efficiency of the verbal and mathematical sections of the *Scholastic Aptitude Test*.

The *Guilford-Shneidman-Zimmerman Interest Survey* (53) was developed to provide a "hobby" and "vocation" interest score in 18 special-interest traits within nine general-interest categories. Clark (24) released preliminary work on the development of an interest inventory for the skilled trades, an area which has long been neglected. Keys were constructed for plasterer, milk wagon driver, printer, electrician, painter, baker, sheet-metal worker, and plumber.

The *Sims SCI Occupational Rating Scale* (89) was developed for measuring the social-class identification of individuals. The rationale for this scale and some preliminary research concerning its validity were described by Sims (90).

### New Scales for Existing Inventories

Considerable attention was given to the development of new scales for existing inventories, largely thru item-analysis technics and in some cases without attempts at theoretical support.

Winne (116) developed a neuroticism scale for the MMPI, and Williams (114) continued research on a caudality scale for this inventory. An Ac (Achievement Drive) key for the MMPI was constructed by Gough (47) from item analysis of MMPI responses of two samples of 27 high-school seniors differing in honor-point ratio but matched for intelligence and adjustment. When included with the Otis test and the *Cooperative English Tests*, scores from the Ac key (based on responses to 34 discriminating items) raised the multiple correlation of these tests with three-year honor-point ratio. This validation was carried out in the original full sample of 231 students from which the 54 used in the item analysis were selected, but the scale was also tried out with other groups.

Using 28 items from the MMPI and 32 original items, Gough, McClosky, and Meehl (49) developed a scale for dominance and reported correlations approximating .62 between this scale and group ratings of dominance in a high-school and a college sample.

Strong (98) developed a new key for scoring the interests of Senior Certified Public Accountants. Music teacher keys for both Strong inventories were developed by Kleist, Rittenhouse, and Farnsworth (63), and a 1948 Psychologist key, now being used in scoring all blanks sent to Stanford, was developed by Kriedt (64) who also developed keys for experimental, clinical, guidance, and industrial psychologists.

### Administration, Scoring and Reporting

Stone and Kriedt's (93) modified directions for administering the *Strong Vocational Interest Blank* when used with the Hanks answer sheet resulted in fewer recording errors. A window-stencil method for hand-scoring this inventory was developed by Greene, Osborne, and Sanders (52). Layton (66) developed an IBM card profile to facilitate reporting of results of large scale testing.

### Norms and Reliability

Hanna and Barnette (56) and MacPhail (70) reported *Kuder Preference Record-Vocational* norms for relatively large groups of male veterans. While both studies reported significant differences between obtained and published norms, the scales on which differences occurred and the direction of the differences in the two studies were not systematic. Kuder norms were given for university business-school seniors by Shaffer (87) and for sales trainees by Eimicke (36).

Strong (96) gave information about norms for his *Vocational Interest Blanks*. He also reported high test-retest correlations between scores on his test over periods of time ranging from several weeks to 22 years (97).

The median test-retest correlations were of the same order of magnitude in subjects originally tested when 19 years old as they were in subjects originally tested when they were 32 years old, and only a slight decrease in correlation, if any, occurred as the time between administrations increased.

Norms for twelve occupational groups on the Lee-Thorpe *Occupational Interest Inventory* were provided by MacPhail and Thompson (71), and Daniels and Hunter (31) gave MMPI profiles for 25 occupational groups, 14 of which, however, had fewer than 10 cases in them.

*Bell Adjustment Inventory* norms for 1123 high-school students were provided by Taylor and Capwell (101).

Consideration was given to adequacy of MMPI norms for college groups and to reliability and equivalence of various forms of this inventory. From their investigation of performance of college students on group and individual forms, Gilliland and Colgin (43) concluded that published MMPI norms were too high for such groups and that for 89 advanced students in psychology test-retest and split-half reliability coefficients were not very high. Dobson and Stone (33), using the shortened booklet form with relatively large groups of college freshmen, found scores for local males higher than published norms on eight scales and also found significant sex difference on three scales.

Responses of hospital patients to long and short forms of this inventory were compared by Holzberg and Alessi (60), who found correlations on the order of long-form test-retest reliability coefficients. Macdonald (68) studied responses of college students to shortened group and shortened individual forms on a test-retest basis (one-week interval) and concluded that there was reason to question the validity and reliability of the *shortened* forms. Cottle (25) reported that with the exception of three scales (L, D, Pa), correlation between scores on individual and booklet forms ranged from .72 to .91, for college students, and that for similar groups *full* booklet and individual forms could be used interchangeably.

### Circumvention

Because of the inadequacy of our knowledge concerning the validity of personality and interest inventories in specific situations, circumvention of the intent of these inventories continued to receive attention. Cross (29), Mais (73), and Noll (79) reported that responses to structured inventories could be changed at will. Gough (46) reviewed work on the F minus K dissimulation index for the MMPI and suggested several cutting scores for identifying "fake bad" records.

Green (51) found himself in possession of data that led to the development of methodology for this problem. Green inadvertently had structured inventory responses of two groups of juvenile police officers, one group having completed the inventories for descriptive purposes and the other for selection purposes. He was able to select groups matched on the basis of intelligence and practical judgment. Inventory scores for these groups

were compared. Least circumvention appeared in the *Guilford-Martin Inventory of Factors GAMIN*.

Kuder (65) contributed further to the methodology of this problem in his description of the development of an honesty scale for the *Kuder Preference Record-Personal*. In testing the validity of the scale on a cross-validation sample, he considered joint cutting points for the previously constructed validity scale and the new honesty scale.

## Educational Applications

### Junior High School

The *Bell Adjustment Inventory* and *California Test of Personality* scores of 17 eighth-grade pupils rigid in problem-solving were not found by Cowen and Thompson (27) to differ significantly from those of 17 students flexible in problem-solving. High and low scorers on the *Kuhlmann-Anderson Intelligence Test* were found by Hinkelman (59) to have significantly different scores on the *California Test of Personality*.

### High School

Resnick (84) reported low correlation between personality-test scores and grades in a sample of ninth- and tenth-graders. Gough (48) found correlations of approximately  $-.30$  between number of extracurriculum activities of senior high-school boys and girls and their scores on Drake's introversion-extroversion scale for the MMPI.

### College

*Predictive Validity.* Strong (95) reported high correspondence between the *Vocational Interest Blank* scores of college students and the occupations in which they were engaged 20 years later.

Using the method of multiple discriminant analysis, Bryan (12) analyzed the freshman *Kuder Preference Record-Vocational* scores of college sophomores in five fields of concentration and found that the maximum number of four linear combinations of the nine original scores were necessary to account for the significant variation among the fields.

Pre-entrance MMPI scores and subsequent acceptability as a roommate were found to be essentially uncorrelated by Brody (10). Low relationship between antecedent MMPI scores and rated ability in practice teaching was reported by Michaelis and Tyler (74). Similar low relationships were reported by Hake and Ruedisili (55) between *Kuder Preference Record-Vocational* scores and first semester grades in each of five subjects.

*Status Validity.* Altho Borg (9) reported that scores on both the *Bell Adjustment Inventory* and Strong's artist key were essentially uncorrelated with grade average in a college of arts and crafts, he found some differences among the *Kuder Preference Record-Vocational* profiles of students in three specialties within the art curriculum (7) and differences in the responses of art and nonart students to several of Guilford's inventories (8). Differ-

ences in the responses of art and nonart students on the MMPI were noted by Spiaggia (91).

Self, peer, and expert ratings, and responses to various interest and personality inventories were compared by several investigators. Berdie (4) reported contingency coefficients ranging from .21 to .61 between self-ratings of interest and scores in similar areas of the *Kuder Preference Record-Vocational* and the *Strong Vocational Interest Blank*. Neurotic tendency and sociability scores on the *Bernreuter Personality Inventory* were found by Powell (81) to be essentially uncorrelated with peer and expert ratings. Stanley (92) reported positive relationships between a junior-college student's self-rankings on Spranger's types and on rankings using similar scales of the *Allport-Vernon Study of Values*.

Birge (5) reported that fraternity members with high dominance rating differed from those with low dominance rating in responses to several scales of the *Kuder Preference Record-Personal*. MMPI score differences within various groups of leaders and between leaders and nonleaders were reported by Williamson and Hoyt (115). Political activity leaders evidenced some expected personality differences while fraternity and sorority leaders tended to be "just students." Sherman (88) found that "most emancipated" and "least emancipated" women differed in their responses to the *Bernreuter Personality Inventory*.

*Congruent Validity.* Lough and Green (67) found relatively little correlation between the MMPI and the *Washburne S-A Inventory*. Four *Humm-Wadsworth Temperament Scale* components and four similarly named MMPI scales were found to be essentially uncorrelated in one group by Canning, Harlow, and Regelin (15) and in six groups by Gilliland (42). However, a slight positive correlation between the depression scales of the two inventories was found. Low correlation between MMPI and the *Terman-Miles Attitude-Interest Analysis* masculinity-femininity scales was noted by de Cillis and Orbison (32).

Two groups which differed in adjustment according to MMPI scores were also found to differ in *Kuder Preference Record-Vocational* profiles by Feather (38).

Dressel and Matteson (35) investigated the influence of experience on Kuder scores and found a median correlation of .76 between a subject's scores obtained under standard conditions and scores obtained with directions to answer according to experience rather than interest.

### Professional School

The problem of predicting success in professional schools was treated comprehensively by Stuit (100). Several interest and a few personality measures were considered in this book. In similar studies Glaser (44) found no relationship between pre-entrance MMPI scores and first-year general grade average in a medical school. Weisgerber (110) reported no correlation above .30 when he studied the interrelationship of ratings

of practical nursing success and MMPI scores obtained at the time of rating. On two Guilford inventories, Healy and Borg (57) found profile differences between graduate and student nurses.

### Test Theory

Cureton (30) forcibly drew attention to the pitfalls inherent in a completely empirical approach to test construction. Cureton's illustration of how spurious correlation is achieved when items selected on a sample are rescored was demonstrated by using a fictitious sample, but Kirkpatrick (62) reported a similar finding for some actual data. It is also stimulating to note articles by Travers (106) and Flanagan (39) urging the development of tests within rational hypotheses. This dictum is especially pertinent to construction of personality and interest inventories or keys.

Several new ideas for attitude measurement were tried by Cattell and others (21). Campbell (14) reviewed the literature dealing with indirect assessment of social attitudes and urged more tests of an indirect nature. He defined an indirect measure as one which: (a) the respondents will all strive to do well, (b) is sufficiently difficult or ambiguous to allow individual difference in response, and (c) can be loaded with content relative to the attitude to be measured. This theory seems consistent with Cronbach's (28) finding that response sets in achievement tests become more pronounced as items become difficult or ambiguous.

Gordon (45) investigated the relationship between forced-choice and questionnaire methods of personality measurement. He found consistently higher agreement between nominations and test scores when scores were obtained from forced-choice items rather than questionnaire items.

The work of Guttman on scalogram analysis and of Lazarsfeld on latent structure theory, in a volume of Stouffer (94), is of vital concern to the area of personality and interest measurement. The solution for the latent class model of latent structure analysis which was provided by Green (50) should be noted also.

In a series of articles, Mosteller (76, 77) systematically examined and reconstructed one case of the Thurstone paired-comparison scaling method. This model should not be overlooked.

### Multivariate Analysis and Profile Similarity

The discriminating power of a test or battery of tests has been the concern of many investigations reviewed here. For the most part, the investigators have been content either to report the profiles for the averages of several groups or, at the most, to examine differences in pairs of groups, variable by variable. Except for the study of Bryan, (12) there were no personality- or interest-inventory studies reported during this period in which the test averages for two or more groups were treated as points in an  $n$ -dimensional test space and in which a test was made of whether the points were coincident or not. And this, despite the fact that

Fisher's discriminant function, Mahalanobis' generalized distance, and Hotelling's generalized  $t$ -test have been available for this purpose in the two-group case for a number of years.

During the current period, Bryan (12) independently generalized Fisher's discriminant function so that the technic could be applied to any number of groups. In a recent book, Rao (82) discussed the generalization of Fisher's discriminant function that he achieved prior to Bryan. Rao also provided tests of significance for the multiple discriminant function problem. In the reviewers' opinion, Rao's significance tests are superior to the variance-analysis test proposed by Block, Levine, and McNemar (6) since they will detect all possible conditions of difference in group centroids while the Block, Levine, and McNemar test will not.

Osgood and Suci (80) proposed a statistic that measures the distance of a profile pattern from the profile patterns of all other types. Their proposal is intimately related with Mahalanobis' generalized distance.

Psychologists have been reluctant to accept multiple-discriminant analysis on the grounds that it does nothing that is not accomplished by multiple-regression analysis. Rulon (85, 86) and Tiedeman (104) discussed the differences in these two methods of analysis.

In the event that a test or test battery has discriminating power, the problem of using this information in the interpretation of the test record of an individual arises. Characteristically this problem has been handled in terms of clinical judgment about the proximity of the individual's profile to the profiles of averages for several groups. Coding schemes such as those of Welsh (111), Wiener (113), and Frandsen (40) have been developed in order to simplify judgments of this nature. Other investigators have attempted to refine the judgment by means of coefficients such as the coefficients of profile similarity derived by Cattell (20).

Coding methods and profile similarity coefficients are based upon the geometry of the profile, an erroneous model for problems of this nature. The  $n$  points that are indicated in two dimensional space on a profile are essentially the  $n$  coordinates of a single point in  $n$  space. When test performance is interpreted within the framework of the  $n$ -space model, the problem of the proximity of an individual's test record to the average record for various groups is clarified. It is simply that of determining the proximity of the individual's point to the points for the centroids of several groups. The distance derived by Osgood and Suci provides one type of answer to this problem. The centour score proposed by Tiedeman, Bryan, and Rulon (105) provides another type of answer. A centour score is essentially the centile distance of a point from the centroid for a given group. The centour method of reporting group similarity has the merits of being free from scaling problems encountered in distance methods and of resembling the percentile concepts with which most test interpreters are familiar.

## Bibliography

1. BAEHR, MELANY H. "A Factorial Study of Temperament." *Psychometrika* 17: 107-26; March 1952.
2. BELL, HUGH M. *Personal Preference Inventory*. Palo Alto, Calif.: Pacific Books, 1950.
3. BIRDIE, RALPH F. "Counseling Methods: Diagnostics." *Annual Review of Psychology*. Vol. 1 (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1950. p. 255-66.
4. BIRDIE, RALPH F. "Scores on the Strong Vocational Interest Blank and the Kuder Preference Record in Relation to Self Ratings." *Journal of Applied Psychology* 34: 42-49; February 1950.
5. BIRGE, WILLIAM R. "Preferences and Behavior Ratings of Dominance." *Educational and Psychological Measurement* 10: 392-94; Autumn 1950.
6. BLOCK, JACK; LEVINE, LOUIS; and McNEMAR, QUINN. "Testing for the Existence of Psychometric Patterns." *Journal of Abnormal and Social Psychology* 46: 356-59; July 1951.
7. BORG, WALTER R. "The Interests of Art Students." *Educational and Psychological Measurement* 10: 100-106; Spring 1950.
8. BORG, WALTER R. "Personality Characteristics of a Group of College Art Students." *Journal of Educational Psychology* 43: 149-56; March 1952.
9. BORG, WALTER R. "Some Factors Relating to Art School Success." *Journal of Educational Research* 43: 376-84; January 1950.
10. BRODY, DAVID S. "A Genetic Study of Sociality Patterns of College Women." *Educational and Psychological Measurement* 10: 513-20; Autumn 1950.
11. BROWN, CLARENCE W., and GHISELLI, EDWIN E. "Industrial Psychology." *Annual Review of Psychology*. Vol. 3. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1952. p. 205-32.
12. BRYAN, JOSEPH G. *A Method for the Exact Determination of the Characteristic Equation and Latent Vectors of a Matrix with Applications to the Discriminant Function for More Than Two Groups*. Cambridge, Mass.: Harvard University, Graduate School of Education, 1950. 292 p. (Doctor's thesis)
13. BUROS, OSCAR K., editor. *The Third Mental Measurements Yearbook*. New Brunswick, N. J.: Rutgers University Press, 1949. 1047 p.
14. CAMPBELL, DONALD T. "The Indirect Assessment of Social Attitudes." *Psychological Bulletin* 47: 15-38; January 1950.
15. CANNING, WILLIAM; HARLOW, GEORGE; and REGELIN, CLINTON. "A Study of Two Personality Questionnaires." *Journal of Consulting Psychology* 14: 414-15; October 1950.
16. CATTELL, RAYMOND B. "The Discovery of Ergic Structure in Man in Terms of Common Attitudes." *Journal of Abnormal and Social Psychology* 45: 598-618; October 1950.
17. CATTELL, RAYMOND B. "A Factorization of Tests of Personality Source Traits." *British Journal of Psychology, Statistical Section* 4: 165-78; November 1951.
18. CATTELL, RAYMOND B. "The Main Personality Factors in Questionnaire, Self-Estimate Material." *Journal of Social Psychology* 31: 3-38; February 1950.
19. CATTELL, RAYMOND B. *Personality: A Systematic Theoretical and Factual Study*. New York: McGraw-Hill Book Co., 1950. 689 p.
20. CATTELL, RAYMOND B. " $r_p$  and Other Coefficients of Pattern Similarity." *Psychometrika* 14: 279-98; December 1949.
21. CATTELL, RAYMOND B., and OTHERS. "The Objective Measurement of Dynamic Traits." *Educational and Psychological Measurement* 10: 224-48; Summer 1950.
22. CATTELL, RAYMOND B., and SAUNDERS, DAVID R. "Inter-Relation and Matching of Personality Factors from Behavior Rating, Questionnaire, and Objective Test Data." *Journal of Social Psychology* 31: 243-60; May 1950.
23. CHALLMAN, ROBERT C. "Clinical Methods: Psychodiagnostics." *Annual Review of Psychology*. Vol. 2. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1951. p. 239-58.
24. CLARK, KENNETH E. "A Vocational Interest Test at the Skilled Trades Level." *Journal of Applied Psychology* 33: 291-303; August 1949.
25. COTTLE, WILLIAM C. "Card Versus Booklet Forms of the MMPI." *Journal of Applied Psychology* 34: 255-59; August 1950.

26. COTTLE, WILLIAM C. "A Factorial Study of the Multiphasic, Strong, Kuder, and Bell Inventories Using a Population of Adult Males." *Psychometrika* 15: 25-47; March 1950.
27. COWEN, EMORY L., and THOMPSON, GEORGE C. "Problems Solving Rigidity and Personality Structure." *Journal of Abnormal and Social Psychology* 46: 165-76; April 1951.
28. CRONBACH, LEE J. "Further Evidence on Response Sets and Test Design." *Educational and Psychological Measurement* 10: 3-31; Spring 1950.
29. CROSS, ORRIN H. "A Study of Faking on the Kuder Preference Record." *Educational and Psychological Measurement* 10: 271-77; Summer 1950.
30. CURETON, EDWARD E. "Validity, Reliability, and Baloney." *Educational and Psychological Measurement* 10: 94-96; Spring 1950.
31. DANIELS, EDGAR E., and HUNTER, W. A. "MMPI Personality Patterns for Various Occupations." *Journal of Applied Psychology* 33: 559-65; December 1949.
32. DE CILLIS, OLGA E., and ORRISON, WILLIAM D. "A Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI." *Journal of Applied Psychology* 34: 338-42; October 1951.
33. DOBSON, WILLIAM R., and STONE, D. R. "College Freshman Responses on the Minnesota Multiphasic Personality Inventory." *Journal of Educational Research* 44: 611-18; April 1951.
34. DORCUS, ROY M., and JONES, MARGARET H. *Handbook of Employee Selection*. New York: McGraw-Hill Book Co., 1950. 349 p.
35. DRESSEL, PAUL L., and MATTESON, ROSS W. "The Relationship Between Experience and Interest as Measured by the Kuder Preference Record." *Educational and Psychological Measurement* 12: 109-16; Spring 1952.
36. EIMICKE, VICTOR W. "Kuder Preference Record Norms for Sales Trainees." *Occupations* 28: 5-10; October 1949.
37. EYSENCK, HANS J. "Personality." *Annual Review of Psychology*. Vol. 3. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1952. p. 151-74.
38. FEATHER, DON B. "The Relation of Personality Maladjustments of 503 University of Michigan Students to Their Occupational Interests." *Journal of Social Psychology* 32: 71-78; August 1950.
39. FLANAGAN, JOHN C. "The Use of Comprehensive Rationales in Test Development." *Educational and Psychological Measurement* 11: 151-55; Spring 1951.
40. FRANDSEN, ARDEN N. "A Note on Wiener's Coding of Kuder Preference Record Profiles." *Educational and Psychological Measurement* 12: 137-39; Spring 1952.
41. GILBERT, WILLIAM M. "Counseling: Therapy and Diagnosis." *Annual Review of Psychology*. Vol. 3. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1952. p. 351-80.
42. GILLILAND, A. R. "The Humm-Wadsworth and the Minnesota Multiphasic." *Journal of Consulting Psychology* 15: 457-59; December 1951.
43. GILLILAND, A. R., and COLGIN, RUSSELL. "Norms, Reliability, and Forms of the MMPI." *Journal of Consulting Psychology* 15: 435-48; October 1951.
44. GLASER, ROBERT. "Predicting Achievement in Medical School." *Journal of Applied Psychology* 35: 272-74; August 1951.
45. GORDON, LEONARD V. "Validities of the Forced-Choice and Questionnaire Methods of Personality Measurement." *Journal of Applied Psychology* 35: 407-12; December 1951.
46. GOUGH, HARRISON G. "The F Minus K Dissimulation Index for the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 14: 408-13; October 1950.
47. GOUGH, HARRISON G. "Factors Relating to the Academic Achievement of High School Students." *Journal of Educational Psychology* 40: 65-78; February 1949.
48. GOUGH, HARRISON G. "A Research Note on the MMPI Social I. E. Scale." *Journal of Educational Research* 43: 138-41; October 1949.
49. GOUGH, HARRISON G.; McCLOSKEY, HERBERT; and MEEHL, PAUL E. "A Personality Scale for Dominance." *Journal of Abnormal and Social Psychology* 46: 360-66; July 1951.
50. GREEN, BERT F., JR. "A General Solution for the Latent Class Model of Latent Structure Analysis." *Psychometrika* 16: 151-66; June 1951.

51. GREEN, RUSSEL F. "Does a Selection Situation Induce Testees to Bias Their Answers on Interest and Temperament Tests?" *Educational and Psychological Measurement* 11: 503-15; Autumn 1951.
52. GREENE, JAMES E.; OSBORNE, ROBERT T.; and SANDERS, WILMA B. "A Window-Stencil Method for Scoring the Strong Vocational Interest Blank (Men)." *Journal of Applied Psychology* 33: 141-45; April 1949.
53. GUILFORD, JOY P.; SHNEIDMAN, EDWIN S.; and ZIMMERMAN, WAYNE S. "The Guilford-Shneidman-Zimmerman Interest Survey." *Journal of Consulting Psychology* 13: 302-306; August 1949.
54. GUILFORD, JOY P., and ZIMMERMAN, WAYNE S. *The Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply Co., 1949.
55. HAKE, DOROTHY TERRY, and RUEDISILI, CHESTER H. "Predicting Subject Grades of Liberal Arts Freshmen with the Kuder Preference Record." *Journal of Applied Psychology* 33: 553-58; December 1949.
56. HANNA, JOSEPH V., and BARNETTE, W. LESLIE, JR. "Revised Norms for the Kuder Preference Record for Men." *Occupations* 28: 168-70; December 1949.
57. HEALY, IRENE, and BORG, WALTER R. "Personality Characteristics of Nursing School Students and Graduate Nurses." *Journal of Applied Psychology* 35: 275-80; August 1951.
58. HESTON, JOSEPH C. *Heston Personal Adjustment Inventory*. Yonkers, N. Y.: World Book Co., 1949.
59. HINKELMAN, EMMET A. "Intellectual Level and Personality Adjustment." *Elementary School Journal* 52: 31-35; September 1951.
60. HOLZBERG, JULES D., and ALESSI, SALVATORE. "Reliability of the Shortened Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 13: 288-92; August 1949.
61. HUNT, HOWARD F. "Clinical Methods: Psychodiagnostics." *Annual Review of Psychology*. Vol. 1. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1950. p. 207-20.
62. KIRKPATRICK, JAMES J. "Cross-Validation of a Forced-Choice Personality Inventory." *Journal of Applied Psychology* 35: 413-17; December 1951.
63. KLEIST, MYDELLE; RITTENHOUSE, C. H.; and FARNSWORTH, PAUL R. "Strong Vocational Interest Scales for Music Teachers." *Occupations* 28: 100-101; November 1949.
64. KRIEDT, PHILIP H. "Vocational Interests of Psychologists." *Journal of Applied Psychology* 33: 482-88; October 1949.
65. KUDER, G. FREDERIC. "Identifying the Faker." *Personnel Psychology* 3: 155-67; Summer 1950.
66. LAYTON, WILBUR L. "An IBM Card Profile for the Strong Vocational Interest Blank." *Journal of Applied Psychology* 34: 415-16; December 1950.
67. LOUCH, ORPHA M., and GREEN, MARY E. "Comparison of the Minnesota Multiphasic Personality Inventory and the Washburne S-A Inventory as Measures of Personality of College Women." *Journal of Social Psychology* 32: 23-30; August 1950.
68. MACDONALD, GORDON L. "A Study of the Shortened Group and Individual Forms of the MMPI." *Journal of Clinical Psychology* 8: 309-11; July 1952.
69. MACKINNON, DONALD W. "Personality." *Annual Review of Psychology* Vol. 2. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1951. p. 113-36.
70. MACPHAIL, ANDREW H. "That Changing Kuder." *Occupations* 30: 202-203; December 1951.
71. MACPHAIL, ANDREW H., and THOMPSON, GEORGE R. "Interest Patterns for Certain Occupational Groups: Occupational Interest Inventory (Lee-Thorpe)." *Educational and Psychological Measurement* 12: 79-89; Spring 1952.
72. MACARET, ANN. "Clinical Methods: Psychodiagnostics." *Annual Review of Psychology* Vol. 3. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1952. p. 283-320.
73. MAIS, ROBERT D. "Fakability of the Classification Inventory Scored for Self Confidence." *Journal of Applied Psychology* 35: 172-74; June 1951.
74. MICHAELIS, JOHN U., and TYLER, FRED T. "MMPI and Student Teaching." *Journal of Applied Psychology* 35: 122-24; April 1951.
75. MOONEY, ROSS L., and GORDON, LEONARD V. *Mooney Problem Check Lists, Grades VII-IX, X-XII*. Columbus: Ohio State University Press, 1950.

76. MOSTELLER, FREDERICK. "Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations." *Psychometrika* 16: 3-9; March 1951.
77. MOSTELLER, FREDERICK. "Remarks on the Method of Paired Comparisons: II. The Effect of an Aberrant Standard Deviation When Equal Standard Deviations and Equal Correlations Are Assumed. III. A Test of Significance for Paired Comparisons When Equal Standard Deviations and Equal Correlations Are Assumed." *Psychometrika* 16: 203-18; June 1951.
78. MYERS, R. C., and SCHULTZ, DOUGLAS G. "Predicting Academic Achievement with a New Attitude-Interest Questionnaire—I." *Educational and Psychological Measurement* 10: 654-63; Winter 1950.
79. NOLL, VICTOR H. "Simulation by College Students of a Prescribed Pattern on a Personality Scale." *Educational and Psychological Measurement* 11: 478-88; Autumn 1951.
80. OSGOOD, CHARLES E., and SUCI, GEORGE J. "A Measure of Relation Determined by Both Mean Difference and Profile Information." *Psychological Bulletin* 49: 251-62; May 1952.
81. POWELL, MARGARET G. "Comparisons of Self-Rating, Peer-Ratings, and Expert's Ratings of Personality Adjustment." *Educational and Psychological Measurement* 8: 225-34; Summer 1948.
82. RAO, C. RADHAKRISHNA. *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons, 1952. 390 p.
83. REMMERS, HERMANN H., and SHIMBERG, BENJAMIN. *S. R. A. Youth Inventory, Form A*. Chicago: Science Research Associates, 1949.
84. RESNICK, JOSEPH. "A Study of Some Relationships Between High School Grades and Certain Aspects of Adjustment." *Journal of Educational Research* 44: 321-40; January 1951.
85. RULON, PHILLIP J. "Distinctions Between Discriminant and Regression Analyses and a Geometric Interpretation of the Discriminant Function." *Harvard Educational Review* 21: 80-90; Spring 1951.
86. RULON, PHILLIP J. "The Stanine and the Separile: A Fable." *Personnel Psychology* 4: 99-114; Spring 1951.
87. SHAFFER, ROBERT H. "Kuder Interest Patterns of University Business School Seniors." *Journal of Applied Psychology* 33: 489-93; October 1949.
88. SHERMAN, ARTHUR W., JR. "Personality Factors in the Psychological Weaning of College Women." *Educational and Psychological Measurement* 8: 249-56; Summer 1948.
89. SIMS, VERNER M. *Sims SCI Occupational Rating Scale*. Yonkers, N. Y.: World Book Co., 1952.
90. SIMS, VERNER M. "A Technique for Measuring Social Class Identification." *Educational and Psychological Measurement* 11: 541-48; Winter 1951.
91. SPIAGGIA, MARTIN. "An Investigation of the Personality Traits of Art Students." *Educational and Psychological Measurement* 10: 285-93; Summer 1950.
92. STANLEY, JULIAN C. "Insight into One's Own Values." *Journal of Educational Psychology* 42: 399-408; November 1951.
93. STONE, C. HAROLD, and KRIEDT, PHILIP H. "Modified Directions for Strong Vocational Interest Blank When Used with the Hanks Answer Sheet." *Journal of Applied Psychology* 35: 169-71; June 1951.
94. STOUTER, SAMUEL A., and OTHERS. *Measurement and Prediction*. Studies in Social Psychology in World War II, Vol. 4. Princeton, N. J.: Princeton University Press, 1950. 756 p.
95. STRONG, EDWARD K., JR. "Interest Scores While in College of Occupations Engaged in 20 Years Later." *Educational and Psychological Measurement* 11: 335-48; Autumn 1951.
96. STRONG, EDWARD K., JR. "Norms for Strong's Vocational Interest Tests." *Journal of Applied Psychology* 35: 50-56; February 1951.
97. STRONG, EDWARD K., JR. "Permanence of Interest Scores over 22 Years." *Journal of Applied Psychology* 35: 89-91; April 1951.
98. STRONG, EDWARD K., JR. "Vocational Interests of Accountants." *Journal of Applied Psychology* 33: 474-81; October 1949.
99. STUIT, DEWEY B. "Counseling Methods: Diagnostics." *Annual Review of Psychology*. Vol. 2. (Edited by Calvin P. Stone and Donald W. Taylor.) Stanford, Calif.: Annual Reviews, 1951. p. 305-16.

100. STUIT, DEWEY B., chairman. *Predicting Success in Professional Schools*. American Council on Education Studies. Series VI. Washington, D. C.: the Council, 1949. 187 p.
101. TAYLOR, MAHLON V., JR., and CAPWELL, DORA F. "High School Norms on the Bell Adjustment Inventory, Student Form." *Occupations* 28: 376-80; March 1950.
102. THURSTONE, LOUIS L. "The Dimensions of Temperament." *Psychometrika* 16: 11-20; March 1951.
103. THURSTONE, LOUIS L. *Thurstone Temperament Schedule*. Chicago: Science Research Associates, 1950.
104. TIEDEMAN, DAVID V. "The Utility of the Discriminant Function in Psychological and Guidance Investigations." *Harvard Educational Review* 21:71-80; Spring 1951.
105. TIEDEMAN, DAVID V.; BRYAN, JOSEPH G.; and RULON, PHILLIP J. *The Utility of the Airman Classification Battery for Assignment of Airmen to Eight Air Force Specialties*. Cambridge, Mass.: Educational Research Corp. 1951. 328 p.
106. TRAVERS, ROBERT M. W. "Rational Hypotheses in the Construction of Tests." *Educational and Psychological Measurement* 11: 128-37; Spring 1951.
107. TRAXLER, ARTHUR E., and JACOBS, ROBERT. "Construction and Educational Significance of Structured Inventories in Personality Measurement." *Review of Educational Research* 20: 38-50; February 1950.
108. TYLER, FRED T. "A Factorial Analysis of Fifteen MMPI Scales." *Journal of Consulting Psychology* 15: 451-56; December 1951.
109. VERNAN, PHILIP E. "Classifying High-Grade Occupational Interests." *Journal of Abnormal and Social Psychology* 44: 85-96; January 1949.
110. WEISGERBER, CHARLES A. "The Predictive Value of the Minnesota Multiphasic Personality Inventory with Student Nurses." *Journal of Social Psychology* 33: 3-11; February 1951.
111. WELSH, GEORGE S. "Some Practical Uses of MMPI Profile Coding." *Journal of Consulting Psychology* 15: 82-84; February 1951.
112. WHEELER, WILLIAM M.; LITTLE, KENNETH B.; and LEHNER, GEORGE F. J. "The Internal Structure of the MMPI." *Journal of Consulting Psychology* 15: 134-41; April 1951.
113. WIENER, DANIEL N. "Empirical Occupational Groupings of Kuder Preference Record Profiles." *Educational and Psychological Measurement* 11: 273-79; Summer 1951.
114. WILLIAMS, HAROLD L. "The Development of a Caudality Scale for the MMPI." *Journal of Clinical Psychology* 8: 293-97; July 1952.
115. WILLIAMSON, EDMUND G., and HOYT, DONALD. "Measured Personality Characteristics of Student Leaders." *Educational and Psychological Measurement* 12: 65-78; Spring 1952.
116. WINNE, JOHN F. "A Scale of Neuroticism: An Adaptation of the Minnesota Multiphasic Personality Inventory." *Journal of Clinical Psychology* 7: 117-22; April 1951.
117. WOODMAN, EVERETT M. "Description of a Guidance Instrument Designed To Measure Attitudes Related to Academic Success in College." *Educational and Psychological Measurement* 12: 275-84; Summer 1952.

## CHAPTER V

### Development and Applications of Projective Tests of Personality

JOHN W. M. ROTHNEY and ROBERT A. HEIMANN

**T**HE horns of the dilemma on which an earnest clinician finds himself are clearly seen in the research on projective technics. Frustrated in his attempts to apply the statisticians' generalized procedures and products to the individual case, he turns to the intuitive approach of projective testers and finds little satisfaction there. If he attempts to resolve the issues by undertaking the validation of his projective protocols he is, if he is to be scientifically respectable in the current and perhaps contemporary use of that term, forced to resort to the methods that produced the originally frustrating generalizations.

Much of the research on projective technics is concerned with attempts to escape from the dilemma. There is evidence of awareness of the need for better validation of projective instruments to replace the dogmatic statements and unverified claims of the early workers in this area. There is also, however, a genuine concern about the adequacy of common actuarial methods for the process. Only one author, Stephenson (82), suggested that adequate methodology is available. He claimed that the modern logic of scientific method was on the side of the clinicians rather than the psychometricians.

#### Validity and Reliability Studies

The Rorschach test continued to get most attention in studies of projective technics. Attempts to determine the stability of scores and to discover the relationship between Rorschach responses and other criteria are increasing. The level of sophistication is rising. Gibby (37) showed that scores on "intellectual variables" of the Rorschach are not stable and that changes may be made at will. He suggested that the responses can be interpreted only when the precise conditions of test administration are known and when there is knowledge of the particular population from which the subjects in a sample are drawn. Abramson (1), in one of the better studies of the period, showed that Rorschach results of college students may be altered significantly by set or suggestion. He proposed that the amount of change might be used as a measure of flexibility of normals for comparison with the greater rigidity of pathological cases. Baughman (10) found that Rorschach results may be influenced by the examiner and by differences in scoring procedures. Alden and Benton (4) studied the effect of sex of the examiner on the responses of 50 male and female subjects and found that no differences could be attributed to their influence. Holzberg and Wexler (51) found that 20 chronically ill schizophrenics hos-

pitalized for eight years gave stable Rorschach reports; but Hutt and others (54) found many unstable variables in a nonpsychiatric population. They claimed that the instability of the normal is a capacity to shift—the flexibility of a healthy organism. Carp and Shavzin (22) showed that 20 college students could manipulate their responses to give “good” or “bad” impressions when they took the Rorschach a second time.

Attempts to validate Rorschach findings against case history materials have produced few positive results. Wells' (87) study of Rorschach patterns of 12 Harvard National Scholars led him to conclude that the over-all validity of the Rorschach makes an impression of the same order as similarly competent handwriting analyses. Forer and others (32) conducted a thoro study of 30 Rorschach protocols analyzed by staff psychologists with from three to 10 years experience in the use of the test. The examiners worked out their definitions of signs by elaborate group processes. They found that the inter-rater agreement was low and that group discussion did not increase it. At the end of their study they examined the case folders of their subjects, and confidence in the accuracy of their criteria was shaken. Sacks and Lewin (74) showed the fallibility of Rorschach signs and blind diagnosis in predicting behavior. All of these studies suggested that serious errors could result when projective technics were not supplemented by broader clinical approaches.

Attempts to assess the validity of Rorschach patterns have not produced positive results. Neff and Lidz (63) selected 100 soldiers to reproduce approximately the distribution of intelligence in the wartime army population. He found that the intelligence factor was more important in determining the range and configuration of Rorschach response than had been anticipated. After examination of his data, he suggested that the influence of intelligence on Rorschach responses needs to be re-evaluated. Altus and Thompson (5) administered the group Rorschach, Altus' *Measure of Verbal Aptitude*, and the *Ohio State Psychological Examination* to 228 college students. They reported that the relationship between movement signs in the Rorschach and *Ohio State Psychological Examination* scores was nonlinear ( $\eta$  .54 to .63). Cronbach (28) found that no Rorschach indicators of 200 students at the University of Chicago correlated significantly with total or part scores on the *ACE Psychological Examination*. Anderson (8) found some relationship between group Rorschach scores, supervisors ratings, intelligence, and mechanical aptitude test scores of 86 machinists; but Kates (56) found no significant relationship between Rorschach, *Strong Vocational Interest Blank*, and job satisfaction responses of 100 government clerical workers. Holtzman (50) found that for 46 normal to superior college students, the commonly claimed relationship between Rorschach test data and the personality traits of shyness and gregariousness, as rated by associates, was not supported. Levy (60) measured palmar skin resistance and administered the Rorschach to 50 male college students. She found that there were no statistical differences in galvanic response among the cards used and inferred there was no

affective difference. This study is based on the assumption that palmar skin resistance is a reliable measure of affective behavior, at best a questionable concept.

Shappenfield and Bucker (75), by showing the last three cards of the group Rorschach in black and white and then in color to 238 college students, raised some doubt about the meaning of interpretations based on color. Hamlin, Albee, and Leland (41) found that only 6 of 26 signs distinguished between groups of 20 normal college students, maladjusted persons, and neuropsychiatric Veterans Administration patients. Carp (23) tested the entire third grade, 47 boys and 46 girls, in a public school with the Rorschach. She studied the relationship between scores on that test and performances on *Draw-Your-Own Family*, *Draw-How-You-Feel* tests, and scores on the *McFarland Trait Rating Blank*. Her attempt to get agreement of "constriction" by this process suggested that this trait was specific to the instrument used.

Two studies of the Rorschach by Wittenborn (91, 93) stand out from the others in their design and use of statistical methods. In one study, Wittenborn (93) used the responses of 247 college students to the Rorschach cards. He rejected the usual abstract scoring procedures and set up two statistically testable hypotheses: (a) that all responses falling in a given category are similar in some behavioral aspect; and, (b) that the psychological significance of responses falling in a given category is different in some respect from responses not placed in this category. Both hypotheses failed to be sufficiently supported. In a second study Wittenborn (91), after making a factor analysis of intercorrelations of 21 basic scores obtained by the Klopfer scoring system, demonstrated that four factors and several clustering tendencies could be observed. He concluded that incorrect emphasis may have influenced the development of current Rorschach scoring procedures and interpretative practices. If one is willing to permit the manipulation of Rorschach scores by common statistical procedures, these studies by Wittenborn are convincing. There still remains, however, the question concerning the application of such methods to these kinds of data.

Some of the studies of the *Thematic Apperception Test* showed a higher level of experimental sophistication than those of the Rorschach noted above. Those of Wittenborn (92), and Wittenborn and Eron (94), were again outstanding. In one study (92) he used eight selected cards with 100 undergraduate students to test two hypotheses: (a) that there is no tendency for superficially similar response categories to be consistently related; and (b) that response categories are related with each other in a manner consistent with a dynamic interpretation of behavior. His results suggest that there is some reason to believe that consistent use of a personality theory may help the clinician in his interpretation of TAT records. In a second study Wittenborn and Eron (94) analyzed TAT responses of 100 college students and concluded that the emotional tone of the re-

actions of their subjects to TAT cards appeared to be determined by the cards rather than by homogeneous behavioral tones of the students. The outcome of the stories appeared to be independent of the cards and therefore of some value in assessing the affective level of the individual. Hartman (44) studied relationships among 56 categories in TAT responses and personality ratings on a Likert-type personality rating scale of 35 superior teen-aged boys in a detention home. Most of the biserial coefficients were in the .40 to .55 range, but a coefficient of .82 between TAT vocabulary and rating of fluency was found. Ratings of tidiness correlated .38 with criticisms of TAT pictures. Saxe (76) attempted to validate TAT reports by "blind" analysis against criteria of diagnoses set up by psychiatrists who had attempted therapy with 20 children, aged nine to 17, over a four-month period. He concluded that, altho agreement between the two methods of diagnosis was relatively high, the evidence supporting "blind" interpretation of TAT stories was not very strong. Bellak, Levinger, and Lipsky (14) used psychiatrists and students of the TAT to judge two sets of TAT responses of a 16-year-old girl obtained at an eight-month interval. The agreement of the judges about the chronological sequence in this one case prompted the authors to conclude that the TAT might be a useful guide to the understanding of maturational process of adolescents. Bills, Leiman, and Thomas (17) attempted to study the validity of responses of eight third-grade pupils to 10 cards of the TAT. They rated their subjects on the basis of six play-therapy interviews and responses to 10 colored animal pictures. Three of the 24 intercorrelation coefficients, ranging from  $-.09$  to  $+.58$ , were significant at the 1-percent level. They suggested that animal stories and TAT responses revealed the same needs to a small degree. Bills (16) found that school children, aged five to 10, did not respond to TAT cards or 10 colored animal pictures at sufficient length to satisfy a criterion of average story length of 200 words. A study of the assumptions underlying the Negro version of the TAT by Riess, Schwartz, and Cottingham (70) indicated that there was no significant difference in productivity of responses to the Negro form of the test by 30 Negro and 30 white female college students. The authors questioned the hypothesis that the TAT can distinguish between cultural groups.

The validation of some of the lesser known projective technics and some new ones have produced generally negative results. Pascal and Suttell (65) reported their study of the quantification and validity of Bender-Gestalt responses of adults. Using a new scoring system with 40 normals, 40 neurotics, and 40 psychotics they obtained a reliability coefficient of .90. The test-retest coefficients of scores of 23 normals over a period of 18 months was .63, and biserial coefficients between scores derived by the new method and psychiatric diagnoses for 23 normals and psychotics ranged from .76 to .79. Kitay (57) used the responses of 60 college students to work out an objective method of scoring the *Bender-Gestalt Test*. A split-half method of computing reliability, not suitable for the data, produced a coefficient of .75. No evidence of validity was presented. French (36)

used analysis-of-covariance methods for the study of the reactions of 80 college students who had been given false reports on their classroom examination scores and then retested with the *Rosenzweig Picture Frustration Test*. He found that good students who were purposely given lower grades than they had earned did not display more frustration than those who were given their correct grades. The effect of the examiner's personality on subjects' selections of Szondi pictures was shown to be very great in a study by Scherer and others (77). Fosberg (33) found in his testing of 200 subjects that the Szondi pictures did not discriminate between normal and abnormal persons. He showed that altho chance was not the sole determiner in a subject's choices of pictures, the factors which do determine selections are not clear. He indicated that the test should be looked upon with great skepticism and should not be used clinically until some of the basic problems of this instrument are solved. Rotter, Rafferty, and Schachtitz (73) computed correlation coefficients between ratings of adjustment of 206 college men and women by college psychologists and *Rotter Incomplete Sentences Blank* scores. The coefficients were .64 for the college women and .77 for college men. Seaton (79) found that incomplete stories with multiple-choice endings designed as a projective technic did not differentiate between a control group of 280 normal children and an experimental group of 50 children rejected by their parents. Albee and Hamlin (3) administered the *Draw-a-Person Test* to 10 subjects in a Veterans Administration clinic. They used 15 clinical psychologists as judges and found a rank-order correlation coefficient of .62 between clinical diagnoses and "blind" inspection of the subjects' drawing. Staples and Conley (81) studied the finger paintings of three- and four-year-old children. They concluded that the use of finger paintings for personality diagnoses at this level was not justified.

Rosenzweig (71) made a vigorous plea for unified effort to establish validation data for projective technics. He proposed nine steps in clinical validation of old and new tests, including a diagnostic clinic of experts from various schools of thought.

It seems clear from the research reviewed above that the validity and reliability of projective technics have not been satisfactorily established. There is some evidence that the problem of validity has been recognized and that many clinicians realize that even such tests as the Rorschach are still in the earliest stage of validation investigation. There are fewer incantations; fewer statements that blots, pictures, or drawings are mirrors to reflect the mind in a manner unrecognizable except to the projective tester; fewer statements to the effect that projections are immune to statistical treatment. These may, of course, be reflected only by those who write—not by all projective test users. There is, however, much evidence that the designs and methods of researches could be improved. Small sample statistics have encouraged experimental designs that are inadequate and, at times, they *seem* to answer questions that could not be answered without more thoro studies. At times it seems that the clinician needs a wholly

new set of technics applicable to his particular problems. When such methods are devised perhaps the reports of projective testers will resemble experiments more than advertisements.

### Normative Procedures

It is startling to discover in some general discussions of projective technics the admission that separate norms for different groups may be required. To those who are familiar with the norms given in a well-standardized achievement test, a statement, in 1952, by Carlson (21) that the most important finding in a study of Rorschach responses of 100 eighth-grade children is that variability is great and that some deviation of responses from adult norms is to be expected in children's responses, is indicative of the present stage of development in the consideration of projective normative data. It is disconcerting to note that the establishment of norms has been so long delayed, but it is encouraging to find that Ledwith (59) began a longitudinal study of Rorschach responses of a sample of 160 children, ages six to 12, representing one child per thousand in Pittsburgh and Allegheny County. Cass and McReynolds (24) have developed norms, percentiles, means, and sigmas of Rorschach responses of 58 males and 46 females who composed a fairly representative group. The attempt may be less effective than it might be, because some of the tests were administered by graduate students who had given fewer than 20 tests. These two studies represented the beginning of a statistical standardization which their authors claimed had been long overdue. Beck (11) reported more comprehensive norms for adults in a revision of his volume on basic Rorschach processes.

Normative studies for projective technics other than the Rorschach have been reported by several investigators. Rosenzweig (72) provided revised norms for his *Picture Frustration Test* based upon the responses of 236 males and 224 females aged 20 to 29 years. He reported means, standard deviations, frequencies, and percents of responses in various scoring categories. Harriman and Harriman (42) found differences, ascribed to maturation, between performances of 30 children, five to seven years of age, on the *Bender Visual Motor Gestalt Test*. Andrew and others (9) reported some preliminary normative work on a thematic apperception test for children entitled the *Michigan Picture Test*. Ten of their cards were standardized on a random sample of Michigan school children. They reported that much normative data were needed for interpretation of thematic apperception scales. Eron (29) published a table of popular responses of six groups of 150 subjects to the TAT. Eron and Ritter (30) obtained written and oral responses to TAT pictures from groups of 30 college students and suggested that more norms for written responses to the test should be obtained.

Three studies indicate that national and cultural group norms are needed. Stewart and Leland (83) studied the differences on the mosaics made by 128 English and 82 American children. They found significant

differences in the types produced, even to the extent that one type that was thought in England to be an indication of emotional disturbance was made frequently by the most stable American children. Differences in previous training and mental habits between such groups suggested the need for national norms. Buhler (19) found significant differences among 264 Austrian, English, Dutch, Norwegian, and American children in projection patterns in the *World Test*, and Buhler, Lumry, and Carrol (20) summarized studies in the standardization of that technic. Goldenberg (38) published his findings on the responses to the *Make-a-Picture-Story-Test* of seven groups of children, including disturbed adolescents and asthmatic children.

Altho the need for norms in the field of psychometrics is usually well recognized, it has not been so apparent to the users of projective technics, in some cases it almost appears to have been an afterthought. The authors of new tests appeared to be striving to provide norms in a matter not previously common in this area. It should be recognized that the characteristics presumed to be measured by projective technics are not always well defined because the desirability of certain kinds of behavior is not as clearly evident as in the case of achievement or aptitude tests. Nor, since the administration and scoring of projective technics is so time-consuming, is it as easy for the clinician to get large populations as it is for testers in other fields. In view of these limitations, normative procedures for projective technics seem to lag behind those used in the more simple achievement and aptitude testing programs. Much needs to be done.

### Applications of Projective Technics

Projective technics have been used or proposed for use in the study of such groups as obese women, blind adults, stutterers, adoptive parents, discordant marriage partners, children with reading disabilities, hospitalized schizophrenics, persons with suicidal tendencies, Indians within certain cultures, unsuccessful students, and many others. Since space requires some selection from voluminous research, the studies reported below have been chosen as representative of those most likely to be of interest to readers of the REVIEW OF EDUCATIONAL RESEARCH.

Estvan (31) used a combined interview and projective method to study social problem awareness of elementary-school children. Sixty children of upper socio-economic status were paired with 60 lower-status children on the basis of IQ, CA, grade, and sex. Each child was shown one picture of poverty. Initial responses and replies to questions about the picture were recorded and analyzed by competent judges. He found that the projective interview procedure appeared to be well suited for the purposes of examining young children's awareness of social problems. This study is superior in design and execution to most in this area, and further research at this high level is needed.

Johnson (55) used six pictures designed to get at racial attitudes with 90

Spanish-American and 90 American children. Scoring of the responses was more reliable than is common in projective work, and the prejudice score derived from it suggested that the technic had some promise. In a well-designed study, Sewell (80) used a locally constructed, unpublished projective device combined with personality tests to study the personality adjustments and traits of children who had undergone varying training experiences. His results, admittedly requiring further verification, cast serious doubts on the validity of psychoanalytic claims regarding the importance of infant disciplines and the efficacy of prescriptions based on them.

Cronbach (28) found that Rorschach performances were not good statistical predictors of college marks at the University of Chicago. The correlation coefficients between Rorschach patterns and marks of 200 students was low (.25), and the relationship between the projective test results and underachievement was not significant. Coefficients between rated adjustment, reputation questionnaire scores, ratings in dormitory units, and Rorschach scores were .17, .20, and .31. He suggested that altho the Rorschach was not a good statistical predictor, it might help the psychologically-trained counselor to guide students. It was also suggested that analysis of tests and criteria might be more useful than over-all scores. Wittenborn (90) studied the relationship between Rorschach protocols, intelligence-test results, and scores on the *Yale Aptitude* battery made by 68 Yale students. He found no linear relationship of significant size between performances on tests and any one Rorschach category and no evidence that certain types of projective responses were correlated with any type of ability. Osborne, Sanders, and Greene (64) found that the addition of group Rorschach results to *American Council on Education Examination* scores raised the multiple R from .56 to .62 in prediction of grades of 504 college freshmen. Tucker (85) compared the Wechsler-Bellevue and Rorschach scores of 100 randomly selected veterans in New Jersey and found that the relationships were not high enough to be of any predictive value.

Cooper and Lewis (26) administered the Rorschach to teachers who had been rated as best-liked and least-liked by junior and senior high-school students. The overlapping of Rorschach responses was so great that individual prediction of acceptance by pupils was impossible. Biber and Lewis (15) devised a projective picture test to explore the feelings of 94 first- and second-grade school children about their relationships to their teachers. They concluded that it is "possible for a teacher to mold attitudes and values thru the classroom atmosphere she creates." Monroe (62) used pictures of children selected from magazines. She asked school children to pretend that a child in a picture was having difficulty with his school work and to compose a story telling of the child's troubles. It was suggested that this projective method might be used in diagnosis of learning disabilities. Beier, Gorlow, and Stacey (12) indicated, after trying the TAT with 40 mentally defective girls with mean Binet IQ scores of 62,

that projective technics might be useful as entering wedges in the study of the fantasy life of mental defectives.

Hallowell (40) illustrated with materials from Objibwa Indian culture the possibility of using projective methods in studying acculturation. Hayes (45) studied prejudices of 67 graduate students in a teachers college with the *Rosenzweig Picture Frustration Test*. McCary (61) used the same test in his study of white and Negro high-school youth in the North and South. He indicated that definite differences in racial and cultural aggressive reactions to frustrations could be observed, and he believed that these could be modified by age and experience. Reynolds (69) used 20 pictures of heads and asked her subjects to fit bodies to them. The protocols suggested that projections could be used to discover racial attitudes.

Reiger's (67, 68) two studies of the use of the Rorschach in the analysis of occupational personalities and selection of workers indicated that the Rorschach could not be used reliably for selection, placement, or guidance in industry. Two reports of application of projective technics differing from those noted above reflected the continuation in some quarters of the uncritical use of instruments. Buck (18) used the *House-Tree-Person Test* in describing a case of marital discord. The statements of elaborate implications from details of drawings was done without question and without evidence. Vorhaus (86), who feels that the Rorschach "... so often seems to have a wisdom beyond that of its interpreter" indicated that it could be used to study the adjustment potentials of individuals prior to entering each of several new phases in psycho-cultural development. Extremes of impressionism in application constituted a small minority of published research reports, but there was no indication of the extent to which they are used in clinical practice.

### New Instruments

The most prominent of the newer projective devices is the *Children's Apperception Test* described by Bellak and Bellak (13). They suggested that children of ages three to 11 frequently identify more readily with figures of animals than figures of persons. The test consists of 10 plates of pictures of animals and is designed to facilitate the understanding of children's relationships to their most important figures and drives. Samples of the kinds of stories usually elicited were described by the authors. Heppell and Raimy (47) used 50 pictures of parent-child relationships with 30 institutionalized delinquents and suggested that this technic could be used as an aid to the interviewer.

In the task of completing incomplete drawings and symbols, three workers claimed that they found some evidence of projection. Franck and Rosen (34) used 36 incomplete drawings and found sex differences in closure. Men were said to close off stimulus areas and to enlarge and expand the stimuli drawings. Women were reported to leave stimulus areas open and tended to blunt or enclose their drawings with sharp lines. No validation data on these findings were reported. Analysis of completions

of the incomplete drawings of the *Horn-Hellersberg Test* by form, content, and perspective appeared to reveal the individual's relation to reality, according to Hellersberg (46). Krout (58) used completion and naming of abstract visual forms (half-circles and half-ellipses) with 157 white Americans and 12 American Indians as a projective device. Validation was attempted against scores on the *California Personality Test* and responses on other projective technics. The author pointed out the need for further research on this test. Goodenough and Harris (39) reviewed research on children's drawings. Their article may be read with profit by those who propose to use drawings as projective technics.

Following the tautophone method, Hutchins (52) used nonmeaningful verbal structures as a projective device. Subjects were instructed to read stimuli of syllables, nonmeaningful words, and some meaningful words arranged in a series, and were asked to tell stories about them. Reports of the results with five graduate students were reported. Stone (84) published his preliminary work with an auditory apperception test. Recorded sounds of crowds, animals, mechanical devices and others were presented, and subjects were required to tell what happened in the noise-making situation Harrower (43) reported results of having 500 persons undergoing therapy draw the most unpleasant things they could think about. This new five-minute test was not validated, but the author speculated on possible clinical use. Wertheimer and McKinney (88) analyzed responses on preinterview blanks of 200 normal University of Missouri students and 200 psychoneurotic subjects. They counted the words in the subjects' responses, examined the vivid words used, and analyzed the use of the space provided. They reported that their method proved useful. Ammons, Butler, and Herzog (6) developed a new *Vocational Apperception Test* composed of plates representing vocations, 10 for women and eight for men. Trial with 35 college men and 40 college women indicated "reasonably high" validity by comparison with *Strong Vocational Interest Blank* scores and personal information.

### Miscellaneous Discussion and Reports

Two major volumes covering the field of projective technics appeared during the period under review. Anderson and Anderson (7) presented a collection of writings by experts in the field. The first 100 pages of this book on problems in the validation of projective technics were particularly significant since they faced squarely the lack of validation data and indicated that the problem had not even been attacked in a substantial and adequate fashion. The volume by Abt and Bellak (2) contained 14 essays of uneven quality ranging from explanations of Rorschach inspection methods to general articles on such technics as finger painting and figure drawing. There were many hypotheses but few data. Frank (35) described the use of projective technics in the study of the individual and raised many problems on which research is needed. Hertz (48) published a comprehensive discussion of Rorschach theory and technic which con-

tained much sound criticism. Holt (49) provided a valuable supplementary classified bibliography on the TAT.

### Conclusion

Despite the abundant criticisms of projective technics, no one has yet answered Hutt (53), who completed his article on the assessment of individual personalities by projective technics with the question, "Can any test do the job better?" If one can shed biases and look directly at the several methods of studying personality that have been proposed, some of the claims for the projective technics must seem as extreme as those made by factor analysts such as Cattell (25). Cronbach (27) indicated that perhaps 90 percent of the conclusions published as a result of statistical treatment of the Rorschach were not substantiated. He said that they were not necessarily false but were based on unsound analysis, and he suggested that new statistical tools were needed. Rabin (66) also made a plea for better statistical devices to use in the study of the individual. Windle (89) claimed that until better statistical tools in this area are developed, the value of projective technics cannot be determined.

As Schofield (78) has pointed out in a thoro statement, it appears that clinicians are now in the process of trying to separate what has been merely claimed from what has been sufficiently demonstrated. In an area in which the current range is from extremes of objectivity to extremes of impressionism this separation appears to be badly needed, and development of the process constitutes the major trend in this area. In it, however, the clinician still finds himself on the horns of the dilemma stated in the first paragraphs of this review.

### Bibliography

1. ABRAMSON, LEONARD S. "The Influence of Set for Area on the Rorschach Test Results." *Journal of Consulting Psychology* 15: 337-41; August 1951.
2. ABT, LAWRENCE E., and BELLAK, LEOPOLD, editors. *Projective Psychology*. New York: Alfred A. Knopf, 1950. 485 p.
3. ALBEE, GEORGE W., and HAMLIN, ROY M. "An Investigation of the Reliability and Validity of Judgments Inferred from Drawings." *Journal of Clinical Psychology* 5: 389-92; October 1949.
4. ALDEN, PRISCILLA, and BENTON, ARTHUR L. "Relationship of Sex of Examiner to Incidence of Rorschach Responses with Sexual Content." *Journal of Projective Techniques* 15: 231-34; June 1951.
5. ALTUS, WILLIAM D., and THOMPSON, GRACE M. "The Rorschach as a Measure of Intelligence." *Journal of Consulting Psychology* 13: 341-47; October 1949.
6. AMMONS, ROBERT B.; BUTLER, MARGARET N.; and HERZIG, SAM A. *The Vocational Apperception Test: Plates and Manual*. New Orleans: R. B. Ammons, Tulane University, 1949. 25 p.
7. ANDERSON, HAROLD H., and ANDERSON, GLADYS L., editors. *An Introduction to Projective Techniques*. New York: Prentice-Hall, 1951. 720 p.
8. ANDERSON, ROSE G. "Rorschach Tests Results and Efficiency Ratings of Machinists." *Personnel Psychology* 2: 513-24; Winter 1949.
9. ANDREW, GWEN, and OTHERS. "The Michigan Picture Test: The Stimulus Values of the Cards." *Journal of Consulting Psychology* 15: 51-54; February 1951.
10. BAUGHMAN, EMMETT E. "Rorschach Scores as a Function of Examiner Difference." *Journal of Projective Techniques* 15: 243-49; June 1951.

11. BECK, SAMUEL J. *Rorschach's Test. (I.) Basic Processes*. Second revised edition. New York: Grune and Stratton, 1949. 227 p.
12. BEIER, ERNST G.; GORLOW, LEON; and STACEY, CHALMERS L. "The Fantasy Life of the Mental Defective." *American Journal of Mental Deficiency* 55: 582-89; April 1951.
13. BELLAK, LEOPOLD, and BELLAK, SONYA S. *Children's Apperception Test*. New York: Children's Psychiatric Service Co., 1949. 13 p. (10 plates)
14. BELLAK, LEOPOLD; LEVINGER, LEAH; and LIPSKY, ESTHER. "An Adolescent Problem Reflected in the Thematic Apperception Test." *Journal of Clinical Psychology* 6: 295-97; July 1950.
15. BIBER, BARBARA, and LEWIS, CLAUDIA. "An Experimental Study of What Young School Children Expect from Their Teachers." *Genetic Psychology Monographs* 40: 3-96; August 1949.
16. BILLS, ROBERT E. "Animal Pictures for Obtaining Children's Projections." *Journal of Clinical Psychology* 6: 291-93; July 1950.
17. BILLS, ROBERT E.; LEIMAN, CHARLES J.; and THOMAS, RICHARD W. "A Study of the Validity of the Thematic Apperception Test and a Set of Animal Pictures." *Journal of Clinical Psychology* 6: 293-95; July 1950.
18. BUCK, JOHN N. "The Use of the House-Tree-Person Test in a Case of Marital Discord." *Journal of Projective Techniques* 14: 405-34; December 1950.
19. BUHLER, CHARLOTTE. "National Differences in 'World Test' Projection Patterns." *Journal of Projective Techniques* 16: 42-55; March 1952.
20. BUHLER, CHARLOTTE; LUMRY, GAYLE K.; and CARROL, HELEN S. "World Test Standardization Studies." *Journal of Child Psychiatry* 2: 2-81; January 1951.
21. CARLSON, RAE. "A Normative Study of Rorschach Responses of Eight Year Old Children." *Journal of Projective Techniques* 16: 56-65; March 1952.
22. CARP, ABRAHAM L., and SHAVZIN, ARTHUR R. "The Susceptibility to Falsification of the Rorschach Psychodiagnostic Technique." *Journal of Consulting Psychology* 14: 230-33; June 1950.
23. CARP, FRANCES M. "Psychological Constriction on Several Projective Tests." *Journal of Consulting Psychology* 14: 268-75; August 1950.
24. CASS, WILLIAM A. JR., and McREYNOLDS, PAUL W. "A Contribution to Rorschach Norms." *Journal of Consulting Psychology* 15: 178-83; June 1951.
25. CATTELL, RAYMOND B. "On the Disuse and Misuse of P, Qs, and O Techniques in Clinical Psychology." *Journal of Clinical Psychology* 7: 203-14; July 1951.
26. COOPER, JAMES G., and LEWIS, ROLAND B. "Quantitative Rorschach Factors in the Evaluation of Teacher Effectiveness." *Journal of Educational Research* 44: 703-707; May 1951.
27. CRONBACH, LEE J. "Statistical Methods Applied to Rorschach Scores." *Psychological Bulletin* 46: 393-429; September 1949.
28. CRONBACH, LEE J. "Studies of the Group Rorschach in Relation to Success in the College of the University of Chicago." *Journal of Educational Psychology* 41: 65-82; February 1950.
29. ERON, LEONARD D. "A Normative Study of the Thematic Apperception Test." *Psychological Monographs*, No. 315. Washington, D. C.: American Psychological Association, 1950. 48 p.
30. ERON, LEONARD D., and RITTER, ANNE M. "A Comparison of Two Methods of Administration of the Thematic Apperception Test." *Journal of Consulting Psychology* 15: 55-61; February 1951.
31. ESTVAN, FRANK J. "The Relationship of Social Status, Intelligence and Sex of Ten and Eleven Year Old Children to an Awareness of Poverty." *Genetic Psychology Monographs* 46: 3-61; August 1952.
32. FORER, BERTRAM R., and OTHERS. "Consistency and Agreement in the Judgment of Rorschach Signs." *Journal of Projective Techniques* 16: 346-51; September 1952.
33. FOSBERG, IRVING A. "Four Experiments with the Szondi Test." *Journal of Consulting Psychology* 15: 39-44; February 1951.
34. FRANCK, KATE, and ROSEN, EPHRAIM. "A Projective Test of Masculinity-Femininity." *Journal of Consulting Psychology* 13: 247-56; August 1949.
35. FRANK, LAWRENCE K. "Understanding the Individual Through Projective Techniques." *Goals of American Education*. Proceedings of the Fourteenth Educational Conference of the Educational Records Bureau. Washington, D. C.: American Council on Education, 1950. p. 52-62.

36. FRENCH, ROBERT L. "Changes in Performance on the Rosenzweig Picture-Frustration Study Following Experimentally Induced Frustration." *Journal of Consulting Psychology* 14: 111-15; April 1950.
37. GIBBY, ROBERT G. "The Stability of Certain Rorschach Variables under Conditions of Experimentally Induced Sets: I. The Intellectual Variables." *Journal of Projective Techniques* 15: 3-26; March 1951.
38. GOLDENBERG, HERBERT C. "A Resumé of Some Make-a-Picture-Story (MAPS) Test Results." *Journal of Projective Techniques* 15: 79-86; March 1951.
39. GOODENOUGH, FLORENCE L., and HARRIS, DALE B. "Studies in the Psychology of Children's Drawings: 1928-1949." *Psychological Bulletin* 47: 369-433; September 1950.
40. HALLOWELL, A. L. "The Use of Projective Techniques in the Study of the Socio-Psychological Aspects of Acculturation." *Journal of Projective Techniques* 15: 27-44; March 1951.
41. HAMLIN, ROY M.; ALBEE, GEORGE W.; and LELAND, EARL M. "Objective Rorschach 'Signs' for Groups for Normal, Maladjusted, and Neuropsychiatric Subjects." *Journal of Consulting Psychology* 14: 276-82; August 1950.
42. HARRIMAN, MILDRED, and HARRIMAN, PHILLIP L. "The Bender Visual Motor Gestalt Test as a Measure of School Readiness." *Journal of Clinical Psychology* 6: 175-77; April 1950.
43. HARROWER, MOLLY R. "The Most Unpleasant Concept Test." *Journal of Clinical Psychology* 6: 213-33; July 1950.
44. HARTMAN, A. ARTHUR. "An Experimental Examination of the Thematic Apperception Technique in Clinical Diagnosis." *Psychological Monographs*, No. 303. Washington, D. C.: American Psychological Association, 1950. 48 p.
45. HAYES, MARGARET L. "Personality and Cultural Factors in Intergroup Attitudes." *Journal of Educational Research* 43: 122-28; October 1949.
46. HELLERSBERG, ELIZABETH F. *The Individual's Relation to Reality in Our Culture: An Experimental Approach by Means of the Horn-Hellersberg Test*. Springfield, Ill.: C. C. Thomas, 1950. 128 p.
47. HEPPELL, H. KENT, and RAIMY, VICTOR C. "Projective Pictures as Interview Devices." *Journal of Consulting Psychology* 15: 405-11; October 1951.
48. HERTZ, MARGUERITE R. "Current Problems in Rorschach Theory and Technique." *Journal of Projective Techniques* 15: 307-38; September 1951.
49. HOLT, ROBERT R. "TAT Bibliography: Supplement for 1950." *Journal of Projective Techniques* 15: 117-23; March 1951.
50. HOLTZMAN, WAYNE H. "Validation Studies of the Rorschach Test." *Journal of Clinical Psychology* 6: 343-51; October 1950.
51. HOLZBERG, JULES D., and WEXLER, MURRAY. "The Predictability of Schizophrenic Performance on the Rorschach Test." *Journal of Consulting Psychology* 14: 395-99; October 1950.
52. HUTCHINS, LEHMAN C. "Nonmeaningful Verbal Structures Used as Projective Material." *Journal of Consulting Psychology* 13: 412-15; December 1949.
53. HUTT, MAX L. "The Assessment of Individual Personality by Projective Tests: Current Problems." *Journal of Projective Techniques* 15: 388-93; September 1951.
54. HUTT, MAX L., and OTHERS. "The Effect of Varied Experimental 'Sets' upon Rorschach Test Performance." *Journal of Projective Techniques* 14: 181-87; June 1950.
55. JOHNSON, GRANVILLE B., JR. "Experimental Projective Technique for the Analysis of Racial Attitude." *Journal of Educational Psychology* 41: 257-78; May 1950.
56. KATES, SOLIS L. "Rorschach Responses Related to Vocational Interests and Job Satisfaction." *Psychological Monographs*, No. 309. Washington, D. C.: American Psychological Association, 1950. 34 p.
57. KITAY, JULIAN I. "The Bender Gestalt Test as a Projective Technique." *Journal of Clinical Psychology* 6: 170-74; April 1950.
58. KROUT, JOHANNA. "Symbol Elaboration Test: The Reliability and Validity of a New Projective Technique." *Psychological Monographs*, No. 310. Washington, D. C.: American Psychological Association, 1950. 67 p.
59. LEDWITH, NETTIE H. "Rorschach Responses of the Elementary School Child: Progress Report." *Journal of Projective Techniques* 16: 80-85; March 1952.
60. LEVY, JEANNE R. "Changes in the Galvanic Skin Response Accompanying the Rorschach Test." *Journal of Consulting Psychology* 14: 128-33; April 1950.

61. McCARY, JAMES L. "Ethnic and Cultural Reactions to Frustration." *Journal of Personality* 18: 321-26; March 1950.
62. MONROE, RUTH L. "Diagnosis of Learning Disabilities Through a Projective Technique." *Journal of Consulting Psychology* 13: 390-95; December 1949.
63. NEFF, WALTER S., and LIDZ, THEODORE. "Rorschach Pattern of Normal Subjects of Graded Intelligence." *Journal of Projective Techniques* 15: 45-57; March 1951.
64. OSBORNE, R. TRAVIS; SANDERS, WILMA; and GREENE, JAMES. "The Prediction of Academic Success by Means of 'Weighted' Harrower-Rorschach Responses." *Journal of Clinical Psychology* 6: 253-58; July 1950.
65. PASCAL, GERALD R., and SUTTELL, BARBARA J. *The Bender-Gestalt Test: Quantification and Validity for Adults*. New York: Grune and Stratton, 1951. 274 p.
66. RABIN, ALBERT I. "Statistical Problems Involved in Rorschach Patterning." *Journal of Clinical Psychology* 6: 1921; January 1950.
67. REIGER, AUDREY F. "The Rorschach Test in Industrial Selection." *Journal of Applied Psychology* 33: 569-71; December 1949.
68. REIGER, AUDREY F. "The Rorschach Test and Occupational Personalities." *Journal of Applied Psychology* 33: 572-78; December 1949.
69. REYNOLDS, RUTH T. "Racial Attitudes Revealed by a Projective Technique." *Journal of Consulting Psychology* 13: 396-99; December 1949.
70. RIESS, BERNARD F.; SCHWARTZ, EMMANUEL K.; and COTTINGHAM, ALICE. "An Experimental Critique of Assumptions Underlying the Negro Version of the TAT." *Journal of Abnormal and Social Psychology* 45: 700-709; October 1950.
71. ROSENZWEIG, SAUL. "A Method of Validation by Successive Clinical Predictions." *Journal of Abnormal and Social Psychology* 45: 507-509; July 1950.
72. ROSENZWEIG, SAUL. "Revised Norms for the Adult Form of the Rosenzweig Picture-Frustration Study." *Journal of Personality* 18: 344-46; March 1950.
73. ROTTER, JULIAN B.; RAFFERTY, JANET E.; and SCHACHTITZ, EVA. "Validation of the Rotter Incomplete Sentences Blank for College Screening." *Journal of Consulting Psychology* 13: 348-55; October 1949.
74. SACKS, JOSEPH M., and LEWIN, HERBERT S. "Limitations of the Rorschach as Sole Diagnostic Instrument." *Journal of Consulting Psychology* 14: 479-81; December 1950.
75. SAPPENFIELD, BERT R., and BUCKER, SAMUEL L. "Validity of the Rorschach 8-9-10 Percent as an Indicator of Responsiveness to Color." *Journal of Consulting Psychology* 13: 268-71; August 1949.
76. SAXE, CARL H. "A Quantitative Comparison of Psychodiagnostic Formulations from the TAT and Therapeutic Contacts." *Journal of Consulting Psychology* 14: 116-27; April 1950.
77. SCHERER, ISIDOR W., and OTHERS. "An Analysis of Patient-Examiner Interaction with the Szondi Pictures." *Journal of Projective Techniques* 16: 225-37; June 1952.
78. SCHOFIELD, WILLIAM. "Research in Clinical Psychology: 1950." *Journal of Clinical Psychology* 7: 215-21; July 1951.
79. SEATON, JAMES K. "A Projective Experiment Using Incomplete Stories with Multiple Choice Endings." *Genetic Psychology Monographs* 40: 149-228; August 1949.
80. SEWELL, WILLIAM H. "Infant Training and the Personality of the Child." *American Journal of Sociology* 58: 150-59; September 1952.
81. STAPLES, RUTH, and CONLEY, HELEN. "The Use of Color in the Finger Painting of Young Children." *Child Development* 20: 201-12; December 1949.
82. STEPHENSON, WILLIAM. "Q-Methodology and the Projective Techniques." *Journal of Clinical Psychology* 8: 219-29; July 1952.
83. STEWART, URSULA G., and LELAND, LORRAINE A. "American Versus English Mosaics." *Journal of Projective Techniques* 16: 246-48; June 1952.
84. STONE, D. R. "A Recorded Auditory Apperception Test as a New Projective Technique." *Journal of Psychology* 29: 349-53; September 1950.
85. TUCKER, J. E. "Rorschach Human and Other Movement Responses in Relation to Intelligence." *Journal of Consulting Psychology* 14: 283-86; August 1950.
86. VORHAUS, PAULINE G. "The Use of the Rorschach in Preventive Mental Hygiene." *Journal of Projective Techniques* 16: 179-92; June 1952.
87. WELLS, FREDERICK L. "Rorschach and Bernreuter Procedures with Harvard National Scholars in the Grant Study." *Journal of Genetic Psychology* 79: 221-60; December 1951.

88. WERTHEIMER, RITA, and MCKINNEY, FRED. "A Case History Blank as a Projective Technique." *Journal of Consulting Psychology* 16: 49-60; February 1952.
89. WINDLE, CHARLES. "Psychological Tests in Psychopathological Prognosis." *Psychological Bulletin* 49: 451-82; September 1952.
90. WITTENBORN, JOHN R. "Certain Rorschach Response Categories and Mental Abilities." *Journal of Applied Psychology* 33: 330-38; August 1949.
91. WITTENBORN, JOHN R. "A Factor Analysis of Rorschach Scoring Categories." *Journal of Consulting Psychology* 14: 261-67; August 1950.
92. WITTENBORN, JOHN R. "The Implications of Certain Assumptions Involved in the Use of the Thematic Apperception Test." *Journal of Consulting Psychology* 14: 216-25; June 1950.
93. WITTENBORN, JOHN R. "Statistical Tests of Certain Rorschach Assumptions: Analyses of Discrete Responses." *Journal of Consulting Psychology* 13: 257-67; August 1949.
94. WITTENBORN, JOHN R., and ERON, LEONARD D. "An Application of Drive Theory to Thematic Apperception Test Responses." *Journal of Consulting Psychology* 15: 45-50; February 1951.

## CHAPTER VI

### Development and Applications of Tests of Educational Achievement in Schools and Colleges

ERIC F. GARDNER

**T**HIS review covers selected literature on tests of educational achievement appearing since the 1950 review by Findley and Smith (40). An attempt has been made to avoid duplication of previous reviews of measurement in specific subjectmatter fields and of such reviews as that by Thorndike (109) and Ebel (33). Because validation studies and applications of achievement tests often include validation and application of tests of intelligence, aptitude, and personality, some overlap with such topics will be inevitable. Readers are advised also to consult the several other chapters of this issue devoted mainly to such topics.

#### Special Problems in Achievement Testing

Aside from technical problems discussed below, a number of papers have focused attention upon certain broad problems in achievement testing. Among these are: (a) the general evaluation of achievement tests, (b) the responsibilities of test producers and publishers, (c) types of new tests needed, and (d) the practical problems inherent in test administration and use.

The first of these problems, the evaluation of achievement tests, was considered by a panel representing four different emphases. Davis (18), representing the point of view of the test editor, stressed the importance of format and validity, with special emphasis on the nature of the individual items as the most important single element affecting validity. Schwab (98), representing the point of view of the subjectmatter specialist, argued that "a test which is highly valid and at the same time highly useful is not possible." He stressed the view that education would benefit much more from validation studies that are more broadly oriented rather than from studies which treat the test as the only variable. He urged closer cooperation between the test constructor and test consumer. Carroll (12), considering the internal statistics of achievement tests, stressed the importance of homogeneity as a criterion. Various definitions and technics of determining test homogeneity, including factor analysis and Loevinger's index, were examined critically and a new definition proposed. The external statistical relationships of achievement tests as criteria for test evaluation were discussed by Gulliksen (54), who proposed that greater attention be paid to relationships of subsequent relevant achievement, training, practice, or drill in the field and to batteries of aptitude tests. In particular, he stressed the importance of evaluating the relationship of the achievement test to a battery of aptitude tests and gave illustrations from military.

research in which such an evaluation resulted in much needed curriculum changes.

The second problem, that of the responsibility of the test producer and publisher, is receiving increasing attention. A "code of ethics" has been proposed recently which, tho not dealing specifically with achievement tests, does have important implications for such producers (1). The problem as to what information test publishers and testing agencies should provide was discussed by Dressel (27), who proposed a 10-point program for test authors or distributing agencies. He also stressed that the main purpose of achievement testing is not that of grading or ranking but of assisting teachers to get maximum achievement or growth. Betts (9) also argued for "longitudinal norms" to be developed by administering tests at the beginning and end of the year. He further urged the inclusion of both norms and goals in the scale of standard tests so that "the two will not be confused as they so often are at present." In view of the differing goals of schools and teachers, this reviewer fails to see how this suggestion can be implemented.

A third problem concerns areas of "achievement" (or development) in which new tests are needed. Various procedures may be utilized to determine such needs. Factor analyses studies, for example, may serve to identify traits for which new tests are needed as well as to suggest means by which a battery of many tests may be replaced by a few. Among recently reported factor analysis studies of test batteries which have included achievement tests are those by Comrey (14), French (45), and Michael, Zimmerman, and Guilford (77). French (44) elsewhere has reviewed and synthesized the findings of 69 factorial studies of tests in the cognitive area. Another approach is to describe the objectives of education in terms of behavioral outcomes and then check existing tests against such objectives to identify gaps. In order to discover those areas of instruction which most seriously lack appropriate measuring devices, at the elementary level, Educational Testing Service recently solicited from a panel of consultants opinions regarding specific behavioral objectives of elementary education. Not only does a statement of objectives in terms of desired pupil behavior yield suggestions for needed developments in standardized tests, but Lewerenz (70) described the way in which evaluation of the city schools of Los Angeles has been made more effective by such statement of objectives.

In general such analyses, as well as other informed opinion, lead to an emphasis upon the need of tests in addition to the conventional achievement test. Husbands and Shores (61), Mallinson (74), Watt (113), and Wrightstone (115) urged greater attention to traits such as interests, attitudes, critical thinking, personality adaptability, understanding and interpretation, and problem solving. Watt (113) suggested that measurement of appreciation, sensitivity, attitudes, interests and values, and emotional and social adjustment are held up not so much by a lack of technic as by lack of a consistent psychological theory or definition by

which to classify such educational outcomes. Travers (110) urged more careful attention to existing research literature, pointing out that many investigators thru ignorance repeat errors of previous work and make use of inadequate criteria of achievement. He urged that in the construction of new tests existing inadequacies be taken into account.

A final problem facing workers in the measurement field concerns the better utilization of achievement tests. The present reviewer believes that more research should be done regarding practical problems encountered by teachers in the classroom (and by students as well) in their use of both standardized and informal tests. Odom and Miles (84) reported that the oral presentation of true-false tests is superior to visual presentation, especially in the case of poorer students. An exploration of the nature of agreement among readers of essay tests by Torgerson and Green utilizing an inverted factor analysis approach, and a reliability study of "atomistic" versus "wholistic" scoring of English essay tests by Coward was reported by the Educational Testing Service (20) in *Developments*. Lefever (68) urged that formalized achievement testing would be more effective if the classroom teacher were given a more important role in achievement testing in uniform systemwide testing programs. Special biases of teachers, which might be neutralized thru use of achievement tests, have been pointed out in certain studies. That women teachers give higher grades than men and that both give higher grades to girls than boys, altho such differences did not appear in the *Gorman-Schrammel Algebra Test*, has been demonstrated by Carter (13). Dole (22) reported a study on the effectiveness of a program for giving college credits by examination, reaching the conclusion that examinations do identify good students and that it is desirable to use such a system of assigning credits by examination results rather than attendance. Fitch, Drucker, and Norton (41) have again demonstrated the motivating effect of frequent testing. A general consideration of classroom use of tests has been presented by Cook (15), who discussed what the teacher needs to know about measurement, and suggested ways in which knowledge of measurement improves the classroom procedure.

### Technical Problems in Test Development

Technical issues in test development will be considered under three categories: (a) validity and reliability, (b) norms, and (c) scaling methods.

*Validity and Reliability.* Altho contributions to a greater understanding of the validity of measurement instruments are made as a result of all research showing relationship among test performances, and between test performance and behavior, several studies were specifically concerned with validity. Schultz (96) examined the comparability of the scores on three mathematics tests of the College Entrance Examination Board and reported that on the average, scores on the mathematics section of the *Scholastic Aptitude Tests* and *Comprehensive Mathematics Tests* were comparable. Sheldon (100), using the *Progressive Reading Test*, the *Van Wagenen-Dvorak Diagnostic Examination of Silent Reading Abilities*, ob-

tained statistically significant differences between criterion groups of good and poor readers on each instrument. Other writers have considered more general and technical validity issues. Durost (31) raised the question as to procedure in a situation where a test has face validity but has been shown statistically to be too difficult for the intended population. Cronbach and Warrington (17) pointed out that for items of the type ordinarily used in psychological tests, the test with uniform item difficulty gives greater overall validity and superior validity for most cutting scores, compared to a test with a range of item difficulties. A new descriptive parameter for tests, the standard length, is defined and related to reliability, correlation, and validity by means of simplified versions of known formulas by Woodbury (114). The amount of information in a test, in the sense of R. A. Fisher, is related to the standard length. A simplified method has been developed by Horst (60) for estimating the minimum validity which a new measure must possess if it is to afford a specified increase in the predictive efficiency of a test battery, while Goheen and Davidoff (51) have presented a graphical method for the rapid calculation of biserial and point biserial correlation in test research. Some aspects of the problem of differential prediction were considered by Mollenkopf (79), who presented formulas for differential prediction and discussed the desirable correlational relationships among predictors and criterion.

A number of studies were concerned with the problem of reliability and related statistics. Dudek (28) discussed the problems of types of errors which are not "tolerated" in developing reliability formulas—i.e., changes in the ability or traits within the individual and the effect these errors have on the reliability coefficient as estimated from the Spearman-Brown formula. Stanley (106) presented a simplified method for estimating the split-half reliability coefficient of a test. It combines the utilization of Rulon's formula for the reliability coefficient of a whole test secured by split-halves, together with Jenkins' short-cut method for computing a standard deviation. Gulliksen (55) presented several methods for estimating the reliability of a partially speeded test without using parallel forms and illustrated the effect of the formula by means of empirical data. Hamilton (56) presented a formula estimating "real" scores from raw scores on a multiple-choice test. Johnson (62) cited evidence to show that specificity or lack of equivalence in comparable forms of a test tends to lower the reliability but does not lower intertrait correlation coefficients. Lord (72), examining the relation of reliability of multiple-choice tests to the distribution of item difficulties, derived an expression in terms of item difficulties and intercorrelations for the curvilinear correlation of test scores on the "ability underlying the test." This ability is defined as the common factor of item tetrachoric correlation coefficients, corrected for guessing. Green (53) presented a procedure for testing whether there is a statistically significant difference between standard errors of measurement of a test obtained from two different groups of subjects.

*Norms.* Recent emphasis has focused attention on the importance of the

selection of appropriate populations for normative purposes. Claims have been made that normative groups should be homogeneous with respect to such variables as geographical location, sex, socio-economic status, and race. Several studies were reported which indicated that the demand by experts for all types of specialized norms may be overemphasized. Thorndike (108), using *Metropolitan Achievement Test* data and data from the 1940 census, studied community variables as predictors of intelligence and academic achievement. As explanations of the low correlations obtained, he suggested that possibly less emphasis was placed on the more conventional skills in better communities, and hence such variables as school expenditures, school salaries, and library facilities might possibly prove better predictor variables. As an alternative hypothesis he suggested that education may be well standardized and that educational achievement is a leveling factor among communities. Lennon (69) reported a study concerning the relationship between intelligence and achievement test results for a group of communities. He concluded that "in Grades II thru V, at least, the relationships between the intelligence and the achievement levels of a community, with a single exception of those for reading, are not sufficiently large to warrant the establishment of differential norms for school systems of varying average intelligence levels."

Ferrell (39) reported a comparative study of sex differences in the school achievement of white and Negro children. No large sex differences among whites or Negroes were revealed in either arithmetic, social studies, or science. In language usage, girls were superior in both groups. White and Negro boys were more variable than girls in all tests. Bullock (11) reported a study on the comparison of academic achievement of white and Negro high-school graduates. For all comparisons, the Negro group was reported as falling well below the white in achievement. The differential was ascribed to difference in expenditure for the two groups, differences in length of school terms, and salary of teacher differential.

Among the studies stressing a more restricted population was one by Dyer (32), who reported a study on the effects of recency of training on the College Board French scores. The *College Entrance Examination* French scores at Harvard were examined for differences which might be attributed to recency of study of the language. Dyer suggested that recency of study should be included in the future for choosing groups for scaling purposes. Spache (104) attempted to reduce various types of norms given for several oral reading tests to a common denominator.

*Scaling Methods.* As a result of the current interest in scaling problems, a number of symposia and articles, many of which have been previously reviewed, have appeared during the past five years. The most recent discussions took place at the 1952 American Psychological Association meetings and at the 1952 Educational Testing Service Invitational Conference on Testing Problems. Among new scaling procedures published is a method for obtaining scale values determined by the method of successive intervals presented by Edwards and Thurstone (37). Gardner (48) reviewed various

types of scales and stressed the need for a scale giving equal intervals. A technic for obtaining an interval scale in terms of K-units was described. The method involves fitting Pearson Type III Curves to overlapping grade-frequency distributions in a trait in such a way that the proportion of cases in each grade exceeding each raw score is the same as that found in the original data.

### **Achievement Tests in the Evaluation of School Methods and Policies**

Basic information relating to the validity of achievement tests is, of course, to be found in evidence indicating the degree to which they are sensitive to differences in achievement, presumably due to improved instructional methods or to various school policies. It is in studies of such matters that achievement tests find a most significant research use. Papers summarized in this section range from reviews regarding the success of such general educational approaches as "progressive education" to studies concerning the success of quite specific methods.

Harding (57) presented a summary of research comparing progressive versus traditional methods of teaching, both in the specific fields of reading, writing, spelling, and arithmetic as well as in general teaching methods, and appeared to conclude in favor of "progressive" methods. Anderson (2) also summarized literature and argued the case for progressive education. An important new emphasis in the assessment of educational outcomes is to be found in two studies by Furst (46, 47), who emphasizes not so much the *specific* outcomes of *specific* methods as the effect of the *organization* of learning experiences upon the *organization* of learning outcomes. This is indeed a difficult problem, tho the importance of the emphasis is obvious. Organization of learning is defined in terms of the degree of intercorrelation of the various tests outcomes. A group from college and a group from public high schools matched on scholastic aptitude, but with the college group showing superiority on achievement measures, were tested in 1945 and again in 1947. The two groups took approximately the same courses during the two-year study. The initial pattern of intercorrelation for the two groups differed, but in both groups there was a small statistically significant increase in correlation over the two-year period. A Holzinger bifactor analysis was also done. In general, it seemed that the technic used in the college did not produce the desired organization to a greater extent than the technic used in high school. The lack of clear-cut results should not discourage further attacks on this problem, perhaps with other methods.

A group of recent studies represents attempts to evaluate general educational outcomes since they concern general assessments of broad groups or gains made over a number of years of education at some level or another. Anderson (3) reported a study which summarized the relative achievement of the objectives of secondary-school science in a sample of 56 Minnesota schools. Moser and Muirhead (80) studied last school grade

completed by military enlisted men as a factor in their performance on the *Tests of General Educational Development and American History*. Silvey (101) reported a study in changes in test scores of students who were tested again as sophomores on part of the freshman battery. Gains were shown on the *American Council of Education Psychological Examination* and the *Nelson-Denny Reading Test*. Heston (59) administered the *Graduate Record Examination* to women of DePauw University when they were sophomores and again when they were seniors. The difference in the means of the two tests were significant for all but the political science majors. Downie (24) discussed some of the problems in general education suggested by a study of the achievement and opinions of a group of college students. An interesting finding indicated that seniors scored no higher than sophomores on the *Cooperative General Culture Test*.

A number of miscellaneous studies concern the effects of particular methods upon particular types of educational outcomes. Gray (52) has summarized 94 investigations of reading conducted during 1950-51. Raths and Rothman (91) reported findings on the effectiveness of teaching the Three R's from studies carried out over the past 30 years. Jones (63) reported greater gains for an experimental group of third-graders in silent reading achievement when given speech training. McGinnis (73) and Robinson (94) reported favorable outcomes for an experimental reading program. Barbe (7) reported a small controlled group study of the outcomes of remedial instruction in which a significant gain was found. Bradley (10) discussed the problem of literacy in the selection of military personnel and pointed out the effectiveness of the special training unit in reducing illiteracy in a short period of time. Glock (50) studied the effect upon eye movement and reading rate of three methods of training, concluding that there was no evidence that technic designed specifically to train eye movement are generally more effective than a technic involving no mechanical control. Baar (5) made an evaluation of enrichment methods of teaching high-school science to ninth-grade students in a New York City junior high school. Smith and Dunbar (102) reported a study on the difference between discussion participants and nonparticipants who had been matched individually for initial test score on *Watson Glaser Test of Critical Thinking*, but found no statistically significant difference between the groups.

In concluding this section, attention is directed to certain studies relating to such general variables as school policy, organization, class-size and experience of teachers. Russell and Eifert (95) compared the achievement of elementary pupils in single- and double-session schools in a California school system, concluding that children in double sessions are not being given an equal opportunity educationally, either in terms of broadness of curriculum or in terms of achievement in subjects involving equal time spent. Dreier (26) reported a study on the differential achievement of rural graded and ungraded school pupils. The sixth-graders from the graded and ungraded schools did not differ significantly in any of the subjects tested,

but children from graded schools showed superiority in certain subjects at the ninth- and twelfth-grade level. Schunert (97) examined the relationship between mathematical achievement and such factors as the amount of teacher training and experience, social background and educational plans of pupils, class size, and school organization. College policy was considered in one paper by Garret (49), who presented a comprehensive review and bibliography of 194 articles on the opposing theories of restricted selection thru college entrance examinations versus the idea of permitting all to enter a college of broad offerings.

### Predictive Studies Involving Achievement Tests

There have been a number of studies which give evidence regarding the predictive effectiveness of certain achievement tests, but space permits little more than a listing of studies. Bailey (6) studied the relationships among the *California Test of Mental Maturity*, *Stanford Binet*, and the *Progressive Achievement Test*. Shaw (99) examined the relationship between Thurstone primary mental abilities and high-school achievement. The optimum combination of primary abilities accounted for from one-fifth to two-thirds of total variance in achievement scores. Frederiksen and Melville (43) examined the effectiveness of the *Strong Vocational Interest Blank* as a predictive instrument for freshmen engineering students. Olsen (85) checked the validities of law-school admission tests, finding a correlation with first-year grades of .40 and, when combined with prelaw grades, a multiple  $r$  of approximately .52. The validity of law-school achievement tests, when corrected for restriction of range, was found to be .51. Krathwohl and others (65), using a varied test battery, reported a study of the prediction of success in architecture courses. Correlations with over-all grades were in the middle thirties but varied with different predictors for individual courses.

Pierson and Jex (88) reported that the *Cooperative General Achievement Tests* were almost as good as the *Pre-Engineering Inventory* in predicting first-year grades in engineering. The best set of predictors were a combination of high-school grades, total score on the *Cooperative English Test*, and the mathematics score on the *Pre-Engineering Inventory*. Remmers, Elliott, and Gage (92) reported that achievement examinations were better predictors of freshman success at Purdue than were scholastic-aptitude tests, but stressed need for different multiple regression equations for different curriculums. Treumann and Sullivan (112) studied the use of engineering and physical-science aptitude tests as predictors of academic achievement of freshmen students at the University of Wisconsin. The *Engineering and Physical Science Aptitude Test* was the best single indicator of achievement, but when combined with a reading test and the American Council of Education *Psychological Test*, it yielded a multiple correlation coefficient of approximately .53. Lannholm and Schrader (67) summarized and discussed studies pertaining to the prediction of success in graduate school afforded by the *Graduate Record Examinations* from

1937 to early 1951. Phearman (87) studied differences between high-school graduates who went to college and those who did not. The use of tests in the public accounting profession is discussed by Traxler (111).

A number of studies have been concerned with the relationship of reading achievement to later school success. Fay (38) reported a study on the relationship between specific reading skills and selected areas measured by the *Stanford Achievement Test*, finding good readers surpassed poor in six out of 15 comparisons. Results on the *Iowa Silent Reading Test* were compared with those of an objective test on comprehension of United Nations publications by Michaelis and Tyler (78). Readability of UN material was determined by using the Lorge formula, the Flesch, and Dale-Chall formula with inconsistent results. Smith (103) found no relationship between laterality and reading achievement in a group of 9-to-11-year-olds. Preston and Botel (89) compared the relationship of reading skill and other factors to academic achievement of students entering the Wharton School of Finance, University of Pennsylvania. Lanier (66) reported a study contrasting those who continued in high school with "dropouts." When the two groups were matched on intelligence, a small difference in reading and arithmetic achievement in favor of those remaining in school was found, but the means were not significantly different.

A group of studies have been concerned with the later school and college success of students differing in important ways in general background. Andrew (4) reported on college success of nonhigh-school graduates. Usually, the *General Educational Development Test of General Mathematics* was found to be less adequate for students who had not graduated from high school than for those who had. Orr (86) compared records made in college by students from fully accredited high schools with records of students having equivalent ability from second- and third-class high schools. Entrants from accredited high schools remained in college longer and more of them returned after absence. There was little difference reported in grade average and honors earned, tho it is to be noted that more of the poorer students from the accredited schools had remained. Frederiksen (42) reported a study on predicting mathematics grades of veteran and nonveteran students, finding that, with a variety of predictive measures, prediction was equally effective for both groups tho nonveteran students in this sample had higher grades.

### **The Relation of Motivational and Personality Factors to Achievement**

It was pointed out above that achievement testers are increasingly aware of the need for "achievement" measures of such nonintellective functions as attitudes, interests, and values. These traits are worthy of measurement in their own right as objectives of education, but they assume importance also as significant variables related to the more conventional subjectmatter goals of education. Certain studies have appeared during the period covered

by this review dealing with this latter problem; they are summarized together at this point.

Among the studies which relate personality factors to achievement is an investigation of motivation as a predictor of college success by DiVesta, Woodruff, and Hertel (21). An orientation inventory was developed which correlated .41 with grades, and when combined with the *Ohio State Psychological Examination* and the revised *Johnson Science Application Test*, gave about as high a multiple  $r$  in predicting first-term grades as did a more extensive battery of aptitude, science, and mathematics tests and regents results. The authors suggest the use of more measures of motivation such as the orientation inventory. However, the general orientation implied by subjectmatter "preference" did not appear important in a study by Dean (19). Several studies have contrasted under- and over-achieving students in an effort to identify motivational and personality factors that might be important in achievement. Dowd (23) reported differences in interests, study habits, sex, and achievement test results between high ability achievers and underachievers among freshmen in the upper 10 percent in ability at the University of New Hampshire. Myers (83) reported 45 out of 148 attitude-interest items discriminated between the over and the underachievers, but concluded that this agreement is actually between stereotype and expressed attitudes.

Several studies have compared the school achievement of groups which might be expected to differ in degree of motivation. Mumma (81) reported no significant differences in achievement between day and residence pupils in a private secondary school. Justman and Forlano (64), after controlling for significant variables, concluded that a group of academic high-school pupils tested were slightly superior to vocational high-school pupils on the *Cooperative Mathematics Test*. Merrell (76) studied the effects of travel, maturity, and essay tests upon the performance of college geography students, reporting that travel experience and previous essay experience were favorably related to achievement, altho no test of significance was given.

With new technological advances, such as radio and television, there is frequently much concern regarding their effects upon school achievement, since the programs presented are likely to have more appeal than does school homework and thus would affect school motivation and achievement. Two studies, one on television and one on radio, are presented here. Dunham (29) reported that altho the average child spent about 30 hours watching recreational television compared with 20 hours on schoolwork, televiewing did not appear to affect achievement. Ricciuti (93) has decried the dulness of radio educational programs and has demonstrated low child interest in them. The test variables revealing the greatest number of reliable differences between radio listeners and non-listeners were IQ and various tests of educational achievement, with the number and location of these differences varying with the program classification.

The relation of special personality factors, such as emotional adjustment, to achievement has not received much direct attention recently, but a few studies of special handicapped groups likely involve such factors to a substantial degree. Sprunt and Finger (105) reported children with auditory deficiency to be inferior to normals in academic achievement. Zintz (116) studied the social and emotional adjustment of handicapped children, reporting that they were approximately six months retarded in educational achievement. Rabin and Geiser (90) reported a study on the achievement of schizophrenics, other psychotics, and nonpsychotics in basic school subjects. All groups followed the pattern of highest level in reading and lowest performance in arithmetic, a finding supposedly characteristic of developmental disorders.

### New Tests and Test Evaluation

Since the development of new tests within subjectmatter fields is discussed in issues of the REVIEW pertaining to those fields, the present summary is concerned primarily with tests developed for research purposes or those utilized in research endeavors. This reviewer, as did Thorndike (109), found relatively few reports on new achievement tests in the research literature of the past three years. Beckman (8) devised a test of mathematical competence, Murray (82) constructed a special test in geometry, Cooper (16) developed a test of Biblical facts, and Sueltz (107) constructed a test to measure mathematical understandings and judgments. A number of these investigators also reported related research.

Attention should be called to several groups of new instruments, reference to which was not found in the journals. Among such tests are new lengthened forms of the *Graduate Record Examination Advanced Tests* (36); a number of special examinations for the various branches of the Department of Defense (34) covering such topics as electrical and radio information and tool relationships, as well as the usual academic subjects; evaluation instruments of the Eight-Year Study (35) developed to measure certain less tangible results of education; the *Essential High School Content Battery* by Harry and Durost (58); the *Evaluation and Adjustment Series* edited by Durost (30); new forms X-2 and Y-2 of the *Iowa Tests of Educational Development* (71); and a new revision of the *Stanford Achievement Test*.

The standard source for evaluative reviews of specific tests and for bibliography regarding tests is Buros' *Mental Measurements Yearbook*. A new edition of this important volume is now in press. Also relevant is a report by Dragositz and McCambridge (25), describing the extent to which colleges have found various types of tests useful.

### Trends and Future Growth in the Development of Educational Tests

Emphasis thru the period of this review continues to be placed on the fact that achievement in subjectmatter areas is only one phase of the

measurement problem. Since the attention of test makers for the past 50 years has been focused on this relatively easier task, the major need and problem is to supplement the reasonably adequate subjectmatter achievement tests with tests which are valid and easily administered in the equally important but more difficult areas of personality, motivation, interests, and other less concrete areas. These problems are discussed in other chapters in this issue.

Considerable attention has been given to the problem of validity and the adequacy of the criterion. The validity and meaningfulness of tests are, of course, determined by the total body of research involving their use. However, it is important to keep in mind the necessity and importance of human judgment in the validation process. Since achievement tests depend so heavily upon face validity, it seems to the reviewer that test makers owe the user a much more adequate description of the area-content sampled by the test. The admonition to "examine the items" for validity can be done effectively by most teachers only when a frame of reference is supplied.

The current interest on scaling, especially the work by Guttman, Lazarsfeld, and Tucker, has tended to support and reinforce the emphasis placed by a number of measurement people on the importance of the individual test item. Since a poor test item cannot be converted into a good one merely by statistical manipulation, any movement, regardless of its other values, which focuses on the basic test unit is making a valuable contribution to the progress of the testing field.

### Bibliography

1. AMERICAN PSYCHOLOGICAL ASSOCIATION, COMMITTEE ON TEST STANDARDS. "Technical Recommendations for Psychological Tests and Diagnostic Techniques: Preliminary Proposal." *American Psychologist* 7: 461-75; August 1952.
2. ANDERSON, ARCHIBALD W. "The Charges Against American Education: What Is the Evidence?" *Progressive Education* 29: 91-105; January 1952.
3. ANDERSON, KENNETH E. "Summary of the Relative Achievements of the Objectives of Secondary School Science in a Representative Sampling of Fifty-Six Minnesota Schools." *Science Education* 33: 323-29; December 1949.
4. ANDREW, DEAN C. "Predicting College Success of Non-High-School Graduates." *School Review* 60: 151-56; March 1952.
5. BAAR, LINCOLN F. "Critical Selection and Evaluation of Enrichment Methods in Junior-High School General Science." *Science Education* 33: 333-43; December 1949.
6. BAILEY, HELEN K. "A Study of the Correlations Between Group Mental Tests, the Stanford-Binet, and the Progressive Achievement Test Used in the Colorado Springs Elementary Schools." *Journal of Educational Research* 43: 93-100; October 1949.
7. BARBE, WALTER B. "The Effectiveness of Work in Remedial Reading at the College Level." *Journal of Educational Psychology* 43: 229-37; April 1952.
8. BECKMANN, MILTON W. "How Mathematically Literate Is the Typical Ninth Grader after Having Completed Either General Mathematics or Algebra?" *School Science and Mathematics* 52: 449-55; June 1952.
9. BETTS, GILBERT L. "Suggestions for a Better Interpretation and Use of Standardized Achievement Tests." *Education* 71: 217-21; December 1950.

10. BRADLEY, GLADYCE H. "A Review of Educational Problems Based on Military Selection and Classification Data in World War II." *Journal of Educational Research* 43: 161-74; November 1949.
11. BULLOCK, HENRY A. "A Comparison of the Academic Achievements of White and Negro High School Graduates." *Journal of Educational Research* 44: 179-92; November 1950.
12. CARROLL, JOHN B. "Criteria for the Evaluation of Achievement Tests: From the Point of View of the Test Editor." *Proceedings of the 1950 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1951. p. 95-99.
13. CARTER, ROBERT S. "How Invalid Are Marks Assigned by Teachers?" *Journal of Educational Psychology* 43: 218-28; April 1952.
14. COMREY, ANDREW L. "A Factorial Study of Achievement in West Point Courses." *Educational and Psychological Measurement* 9: 193-209; Summer 1949.
15. COOK, WALTER W. "What Educational Measurement in the Education of Teachers?" *Journal of Educational Psychology* 41: 339-47; October 1950.
16. COOPER, CLARA C. "Learning of Biblical Facts in College Correlated with Pre-College Learning and Intelligence and General Culture Test Scores." (Abstract) *American Psychologist* 5: 281-82; July 1950.
17. CRONBACH, LEE J., and WARRINGTON, WILLARD G. "Efficiency of Multiple Choice Tests as a Function of Spread of Item Difficulties." *Psychometrika* 17: 127-47; June 1952.
18. DAVIS, FREDERICK B. "Criteria for the Evaluation of Achievement Tests: From the Point of View of the Test Editor." *Proceedings of the 1950 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1951. p. 73-81.
19. DEAN, STUART E. "Relation of Children's Subject Preferences to Their Achievement." *Elementary School Journal* 51: 89-92; October 1950.
20. DIEDERICH, PAUL B., editor. *Developments*. No. 1. Princeton, N. J.: Educational Testing Service, October 1951. p. 2.
21. DIVESTA, FRANCIS J.; WOODRUFF, ASAHEL D.; and HERTEL, JOHN P. "Motivation as a Predictor of College Success." *Educational and Psychological Measurement* 9: 339-48; Autumn 1949.
22. DOLE, ARTHUR A. "Evidence of the Effectiveness of a Program for Giving College Credits by Examination." *Educational and Psychological Measurement* 11: 387-95; Autumn 1951.
23. DOWD, ROBERT J. "Underachieving Students of High Capacity." *Journal of Higher Education* 23: 327-30; June 1952.
24. DOWNIE, NORVILLE M.; PACE, C. R.; and TROYER, M. E. "Problems in General Education Suggested by a Study of the Achievement and the Opinions of Syracuse University Students." *Educational and Psychological Measurement* 11: 76-80; Spring 1951.
25. DRAGOSITZ, ANNA, and McCAMBRIDGE, BARBARA. "Types of Tests and Their Uses in College Testing Programs." *American Psychologist* 7: 299-300; July 1952.
26. DREIER, WILLIAM H. "The Differential Achievement of Rural Graded and Ungraded School Pupils." *Journal of Educational Research* 43: 175-86; November 1949.
27. DRESSSEL, PAUL L. "Information Which Should Be Provided by Test Publishers and Testing Agencies on the Validity and Use of Their Tests: Achievement Tests." *Proceedings of the 1949 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1950. p. 69-74.
28. DUKE, FRANK J. "Concerning 'Reliability' of Tests." *Educational and Psychological Measurement* 12: 293-99; Summer 1952.
29. DUNHAM, FRANKLIN. "Effect of Television on School Achievement of Children." *School Life* 34: 88-94; March 1952.
30. DUROST, WALTER N., editor. *Evaluation and Adjustment Series*. Yonkers, N. Y.: World Book Co., 1952.
31. DUROST, WALTER N. "Issues in the Measurement of Literature Acquaintance at the Secondary School Level." *Journal of Educational Psychology* 43: 31-44; January 1952.
32. DYER, HENRY S. "The Effect of Recency of Training on the College Board French Scores." *School and Society* 70: 105-106; August 1949.

33. EBEL, ROBERT L. "Construction and Validation of Educational Tests." *Review of Educational Research* 20: 87-97; February 1950.
34. EDUCATIONAL TESTING SERVICE. *Annual Report to the Board of Trustees 1950-51*. Princeton, N. J.: Educational Testing Service, 1952. 105 p.
35. EDUCATIONAL TESTING SERVICE. *Evaluation Instruments of the Eight-Year Study*. General Education Series. Princeton, N. J.: the Service, 1942.
36. EDUCATIONAL TESTING SERVICE. *Graduate Record Examination Advanced Tests*. Princeton, N. J.: the Service, 1952.
37. EDWARDS, ALLEN L., and THURSTONE, LOUIS L. "An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals." *Psychometrika* 17: 169-80; June 1952.
38. FAY, LEO C. "The Relationship Between Specific Reading Skills and Selected Areas of Sixth Grade Achievement." *Journal of Educational Research* 43: 541-47; March 1950.
39. FERRELL, GUY V. "Comparative Study of Sex Differences in School Achievement of White and Negro Children." *Journal of Educational Research* 43: 116-21; October 1949.
40. FINDLEY, WARREN G., and SMITH, ALLEN B. "Measurement of Educational Achievement in Schools." *Review of Educational Research* 20: 63-75; February 1950.
41. FITCH, MILDRED L.; DRUCKER, ARTHUR J.; and NORTON, J. A., JR. "Frequent Testing as a Motivating Factor in Large Lecture Classes." *Journal of Educational Psychology* 42: 1-20; January 1951.
42. FREDRIKSEN, NORMAN. "Predicting Mathematics Grades of Veterans and Non-veteran Students." *Educational and Psychological Measurement* 9: 73-88; Spring 1949.
43. FREDRIKSEN, NORMAN, and MELVILLE, STANLEY D. "Improving the Predictive Value of an Interest Test." (Abstract) *American Psychologist* 7: 285-86; July 1952.
44. FRENCH, JOHN W. *Description of Aptitude and Achievement Tests in Terms of Rotated Factors*. Chicago: University of Chicago Press, 1951.
45. FRENCH, JOHN W., and OTHERS. "Factor Analysis of Aptitude and Achievement Entrance Tests and Course Grades at the United States Coast Guard Academy." *Journal of Educational Psychology* 43: 65-80; February 1952.
46. FURST, EDWARD J. "Effect of the Organization of Learning Experiences upon the Organization of Learning Outcomes: I. Study of the Problem by Means of Correlation Analysis." *Journal of Experimental Education* 18: 215-28; March 1950.
47. FURST, EDWARD J. "Effect of the Organization of Learning Experiences upon the Organization of Learning Outcomes: II. Study of the Problem by Means of Factor Analysis." *Journal of Experimental Education* 18: 343-52; June 1950.
48. GARDNER, ERIC F. "Comments on Selected Scaling Technics with a Description of a New Type of Scale." *Journal of Clinical Psychology* 6: 38-44; January 1950.
49. GARRET, HARLEY F. "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Science and Teachers Colleges." *Journal of Experimental Education* 18: 91-138; December 1949.
50. GLOCK, MARVIN D. "The Effect upon Eye-Movements and Reading Rate at the College Level of Three Methods of Training." *Journal of Educational Psychology* 40: 93-106; February 1949.
51. GOHEEN, HOWARD W., and DAVIDOFF, MELVIN D. "A Graphical Method for the Rapid Calculation of Biserial and Point Biserial Correlation in Test Research." *Psychometrika* 16: 239-42; March 1951.
52. GRAY, WILLIAM S. "Summary of Reading Investigations July 1, 1950, to June 30, 1951." *Journal of Educational Research* 45: 401-37; February 1952.
53. GREEN, BERT F., JR. "A Test of the Equality of Standard Errors of Measurement." *Psychometrika* 15: 251-57; September 1950.
54. GULLIKSEN, HAROLD O. "Criteria for the Evaluation of Achievement Tests: From the Point of View of Their External Statistical Relationships." *Proceedings of the 1950 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1951. p. 100-103.
55. GULLIKSEN, HAROLD O. "The Reliability of Speeded Tests." *Psychometrika* 15: 259-71; September 1950.

56. HAMILTON, C. HORACE. "Bias and Error in Multiple-Choice Tests." *Psychometrika* 15: 151-67; June 1950.
57. HARDING, LOWRY W. "How Well Are Schools Now Teaching the Basic Skills?" *Progressive Education* 29: 7-14; October 1951.
58. HARRY, DAVID P., and DUROST, WALTER N. *Essential High-School Content Battery*. Yonkers, N. Y.: World Book Co., 1952. 28 p.
59. HESTON, JOSEPH C. "Educational Growth as Shown by Retests on the Graduate Record Examination." *Educational and Psychological Measurement* 10: 367-70; Autumn 1950.
60. HORST, PAUL. "The Relationship Between the Validity of a Single Test and Its Contribution to the Predictive Efficiency of a Test Battery." *Psychometrika* 16: 57-66; March 1951.
61. HUSBANDS, KENNETH L., and SHORES, J. HARLAN. "Measurement of Reading for Problem Solving: A Critical Review of the Literature." *Journal of Educational Research* 43: 543-65; February 1950.
62. JOHNSON, HELMER G. "Test Reliability and Correction for Attenuation." *Psychometrika* 15: 115-20; June 1950.
63. JONES, MORRIS V. "The Effect of Speech Training on Silent Reading Achievement." *Journal of Speech and Hearing Disorders* 16: 258-63; September 1951.
64. JUSTMAN, JOSEPH, and FORLANO, GEORGE. "Performance of Academic and Vocational High School Pupils on the Cooperative Mathematics Test." *Mathematics Teacher* 45: 267-68; April 1952.
65. KRATHWOHL, DAVID R., and OTHERS. "Predictions of Success in Architecture Courses." (Abstract) *American Psychologist* 7: 288-89; July 1952.
66. LANIER, J. ARMAND. "A Guidance-Faculty Study of Student Withdrawals." *Journal of Educational Research* 43: 205-12; November 1949.
67. LANNHOLM, GERALD V., and SCHRADER, WILLIAM B. "Predicting Graduate School Success: An Evaluation of the Effectiveness of the Graduate Record Examinations." Princeton, N. J.: Educational Testing Service, 1951. 50 p.
68. LEFEVER, D. WELTY. "The Teacher's Role in Evaluating Pupil Achievement." *Education* 71: 203-209; December 1950.
69. LENNON, ROGER T. "The Relation Between Intelligence and Achievement Test Results for a Group of Communities." *Journal of Educational Psychology* 41: 301-308; May 1950.
70. LEWERENZ, ALFRED S. "New Developments in Evaluating Achievement in the Public Schools of Los Angeles." *Education* 71: 237-44; December 1950.
71. LINDQUIST, EVERET F., editor. *Iowa Tests of Educational Development, Forms X-2, Y-2*. Chicago: Science Research Associates, 1952.
72. LORD, FREDERIC M. "The Relation of the Reliability of Multiple-Choice Tests to the Distribution of Item Difficulties." *Psychometrika* 17: 181-94; June 1952.
73. MCGINNIS, DOROTHY J. "Corrective Reading: A Means of Increasing Scholastic Attainment at the College Level." *Journal of Educational Psychology* 42: 165-73; March 1951.
74. MALLINSON, GEORGE G. "The Implications of Recent Research in the Teaching of Science at the Secondary-School Level." *Journal of Educational Research* 43: 321-42; January 1950.
75. MELVILLE, STANLEY D., and FREDRIKSEN, NORMAN. "Achievement of Freshmen Engineering Students and the Strong Vocational Interest Blank." *Journal of Applied Psychology* 36: 169-73; June 1952.
76. MERRELL, RICHARD H. "The Effects of Travel, Maturity, and Essay Tests upon the Performance of College Geography Students." *Journal of Educational Research* 43: 213-20; November 1949.
77. MICHAEL, WILLIAM B.; ZIMMERMAN, WAYNE S.; and GUILFORD, JOY P. "An Investigation of the Nature of the Spatial-Relations and Visualization Factors in Two High School Samples." *Educational and Psychological Measurement* 11: 561-77; Winter 1951.
78. MICHAELIS, JOHN U., and TYLER, FRED T. "Comparison of Reading Ability and Readability." *Journal of Educational Psychology* 42: 491-98; December 1951.
79. MOLLENKOPF, WILLIAM G. "Some Aspects of the Problem of Differential Prediction." *Educational and Psychological Measurement* 12: 39-44; Spring 1952.

80. MOSER, WILBUR E., and MUIRHEAD, JOSEPH V. "School Grade Last Completed by Military Enlisted Men as Factors in Tests of General Educational Development and American History." *Journal of Educational Research* 43: 221-24; November 1949.
81. MUMMA, RICHARD A. "A Comparison of the Achievement of Day and Resident Pupils in a Private Secondary School." *Journal of Educational Research* 44: 99-106; October 1950.
82. MURRAY, JOHN E. "An Analysis of Geometric Ability." *Journal of Psychology* 40: 118-24; February 1949.
83. MYERS, ROBERT C. "The Academic Overachiever: Stereotyped Aspects." *Journal of Experimental Education* 18: 229-38; March 1950.
84. ODOM, CHARLES L., and MILES, RAY W. "Oral Versus Visual Presentation of True-False Achievement Tests in the First Course in Psychology." *Educational and Psychological Measurement* 11: 470-77; Autumn 1951.
85. OLSEN, MARJORIE A. "Validity of the Law School Admission Test for Predicting First-Year Law School Grades." (Abstract) *American Psychologist* 5: 283-84; July 1950.
86. ORR, HARRIET K. "College Achievement. A Comparison of the Records Made in College by Students from Fully Accredited High Schools with Those of Students Having Equivalent Ability, from Second and Third Class High Schools." *Journal of Educational Research* 42: 353-64; January 1949.
87. PHEARMAN, LEO T. "Comparisons of High-School Graduates Who Go to College with Those Who Do Not." *Journal of Educational Psychology* 40: 405-14; November 1949.
88. PIERSON, GEORGE, and JEX, FRANK B. "Using the Cooperative General Achievement Tests To Predict Success in Engineering." *Educational and Psychological Measurement* 11: 397-401; Autumn 1951.
89. PRESTON, RALPH C., and BOTEL, MORTON. "The Relation of Reading Skill and Other Factors to the Academic Achievement of 2048 College Students." *Journal of Experimental Education* 20: 363-71; June 1952.
90. RABIN, ALBERT I., and GEISER, EUGENE. "The Achievement of Schizophrenics, Other Psychotics and Non-Psychotics in Basic School Subjects." *Journal of General Psychology* 41: 125-29; July 1949.
91. RATHS, LOUIS E., and ROTHMAN, PHILIP. "Then and Now: Some Research Findings on Effectiveness of Teaching the Three R's." *Journal of the National Educational Association* 41: 214; April 1952.
92. REMMERS, HERMANN H.; ELLIOTT, DONALD; and GAGE, NATHANIEL. "Curricular Differences in Predicting Scholastic Achievement: Applications to Counseling." *Journal of Educational Psychology* 40: 385-94 November 1949.
93. RICCIUTI, EDWARD A. "Children and Radio: A Study of Listeners and Non-Listeners to Various Types of Radio Programs in Terms of Selected Ability, Attitude, and Behavior Measures." *Genetic Psychology Monographs* 44: 71-143; August 1951.
94. ROBINSON, HARVEY A. "A Note on the Evaluation of College Remedial Reading Courses." *Journal of Educational Psychology* 41: 83-96; February 1950.
95. RUSSELL, DAVID H., and EIFERT, HAROLD J. "A Comparison of Achievement of Pupils in Single- and Double-Session Schools." *California Journal of Elementary Education* 18: 12-16; August 1949.
96. SCHULTZ, DOUGLAS G. "The Comparability of Scores from Three Mathematics Tests of the College Entrance Examination Board." *Psychometrika* 15: 369-84; December 1950.
97. SCHUNERT, JIM. "The Association of Mathematical Achievement with Certain Factors Resident in the Teacher, in the Teaching, in the Pupil and in the School." *Journal of Experimental Education* 19: 219-38; March 1951.
98. SCHWAB, JOSEPH J. "Criteria for the Evaluation of Achievement Tests: From the Point of View of the Subject-Matter Specialist." *Proceedings of the 1950 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1951. p. 82-94.
99. SHAW, DUANE C. "Study of the Relationships Between Thurstone Primary Mental Abilities and High School Achievement." *Journal of Educational Psychology* 40: 239-49; April 1949.
100. SHELDON, WILLIAM D. "Characteristics of the Reading of a Group of Twelfth-Grade Students." *English Journal* 41: 154-55; March 1952.

101. SILVEY, HERBERT M. "Changes in Test Scores after Two Years in College." *Educational and Psychological Measurement* 11: 494-502; Autumn 1951.
102. SMITH, HENRY CLAY, and DUNBAR, DONALD S. "The Personality and Achievement of the Classroom Participant." *Journal of Educational Psychology* 42: 65-82; February 1951.
103. SMITH, LINDA C. "A Study of Laterality Characteristics of Retarded Readers and Reading Achievers." *Journal of Experimental Education* 18: 321-29; June 1950.
104. SPACHE, GEORGE. "A Comparison of Certain Oral Reading Tests." *Journal of Educational Research* 43: 441-52; February 1950.
105. SPRUNT, JULIE W., and FINGER, FRANK W. "Auditory Deficiency and Academic Achievement." *Journal of Speech and Hearing Disorders* 14: 26-32; March 1949.
106. STANLEY, JULIAN C. "A Simplified Method for Estimating the Split-Half Reliability Coefficient of a Test." *Harvard Educational Review* 21: 221-24; Fall 1951.
107. SUELTZ, BEN A. "Mathematical Understandings and Judgments Retained by College Freshmen." *Mathematics Teacher* 44: 13-19; January 1951.
108. THORNDIKE, ROBERT L. "Community Variables as Predictors of Intelligence and Academic Achievement." *Journal of Educational Psychology* 42: 321-38; October 1951.
109. THORNDIKE, ROBERT L. "Tests as Research Instruments." *Review of Educational Research* 21: 450-62; December 1951.
110. TRAVERS, ROBERT M. W. "The Prediction of Achievement." *School and Society* 70: 293-94; November 1949.
111. TRAXLER, ARTHUR E. "Objective Testing in the Field of Accounting." *Educational and Psychological Measurement* 11: 427-39; Autumn 1951.
112. TREUMANN, MILDRED J., and SULLIVAN, BEN A. "Use of the Engineering and Physical Science Aptitude Test as a Predictor of Academic Achievement of Freshman Engineering Students." *Journal of Educational Research* 43: 129-33; October 1949.
113. WATT, REGINALD R. G. "Newer Emphasis in the Construction of Achievement Tests for College Students." *Education* 71: 226-30; December 1950.
114. WOODBURY, MAX A. "On the Standard Length of a Test." *Psychometrika* 16: 103-106; March 1951.
115. WRIGHTSTONE, J. WAYNE. "New Developments in Evaluating Achievement in the High School." *Education* 71: 210-16; December 1950.
116. ZINTZ, MILES V. "Academic Achievement and Social and Emotional Adjustment of Handicapped Children." *Elementary School Journal* 51: 502-507; May 1951.

## CHAPTER VII

### Development and Applications of Tests of Educational Achievement Outside the Schools

JOHN T. DAILEY

THE material to be covered here will include the development and use of educational achievement tests in industry and government and military organizations as well as certain special testing programs such as the *National Teacher Examinations*, the *Graduate Record Examination*, and the *United States Armed Forces Institute Tests of General Educational Development*. Some of the material in this section will be similar to material reviewed by Mollenkopf in Chapter III of this issue of the REVIEW. Studies will be presented which shed light on the relationships between tests of aptitude and educational achievement.

#### Graduate Record Examination

In addition to the usual validation studies, recent studies of the *Graduate Record Examination* have considered the best use to be made of the tests. Lannholm and Schrader (26) reported on studies of the *Graduate Record Examination* at Harvard, Yale, Princeton, Iowa, Michigan, Columbia, and Vanderbilt. It was found that a combination of tests with undergraduate college records produces better prediction than is obtained when college records alone are used. The Advanced Tests in a given field usually take precedence over the Profile Tests in predicting success. Use of the Profile Tests should ordinarily be justified chiefly to identify strengths and weaknesses for guiding student development rather than for predicting over-all success. Jones (24) reported on some of the results of requiring the *Graduate Record Examination* of all seniors at the University of Buffalo. Each student was required to take both the Profile Tests and the Advanced Test in the department of his concentration. The students scored comparatively better on the Advanced Tests than on the Profile Tests. The results suggested that essentially the Profile Tests measure aptitude, while the Advanced Tests are indicative of collegiate effort. Neither test was a valid predictor of graduate grades. The results are quoted as evidence of overspecialization.

#### General Educational Development

Roeber (30) compared the grades at Kansas State Teachers College for a group entered on the basis of the *United States Armed Forces Institute Tests of General Educational Development* and those entering on the basis of high-school graduation. The GED group made poorer grades than the high-school graduates. However, those entered on the GED performed well enough to justify their entry. Wardlaw (38) carried out a questionnaire

survey of GED testing program administrators in 19 states plus members of the Secondary Commission of the North Central Association. The consensus of the groups surveyed was that GED testing conditions should be more rigorously controlled, minimum passing scores should be raised, some high-school attendance should be required, and diplomas on the basis of general educational development should not be awarded at an age earlier than 20 or 21 years. Chausow (4) found a correlation of .65 between GED test grades and grades in a general course in social science. He concluded that the GED tests were of value as diagnostic tests for determining which superior or weak students should receive special attention.

### National Teacher Examinations

Ryans (32) presented the rationale and philosophy behind the development of the *National Teacher Examinations* and their use in the selection of teachers. He frankly admitted the inadequacies of any written test as a primary basis for the selection of teachers but pointed out that properly-constructed tests can provide information about some aspects of teacher qualifications better than any alternative procedures. Such aspects include professional information, mental abilities and basic skills, general cultural background, and subjectmatter knowledge. Ryans (31) performed an analysis of the results of the 1949 testing and found no significant trends as compared with the results of the previous four years. In another study, Ryans (33) compared the results of internal-consistency analysis and validation against an external criterion of teaching behavior ratings on one of the professional information tests of the *National Teacher Examinations*. He found that the two procedures tend to give substantially different results, with the internal-consistency coefficients ranging higher than the external-validity coefficients. All of these studies were hampered by lack of adequate criteria of teacher proficiency. There appears to be a pressing need for the development of such criteria in this field.

### The College Board and the Educational Testing Service

Both of these organizations are primarily concerned with testing in the schools and colleges. However, both have engaged in very extensive testing and research programs for the armed forces. Fuess (17) summarized the World War II research of the Board. During this period the Board became the center of a very extensive contract research program for the armed forces. In addition to major research and testing programs for the selection of officer candidates, the Board engaged in testing research in such diverse fields as radio, electricity, and gunnery. The *1950-51 Annual Report to the Board of Trustees* of the Educational Testing Service (10) outlined its very extensive testing and research programs for nonschool agencies including the various armed forces plus the Veterans Administration, Department of State, Merchant Marine, Coast Guard, Selective Service, and various nongovernmental professional groups. These projects

range from admissions programs for armed forces officer schools and the development of differential classification test batteries to fundamental research on the nature and organization of human skills and aptitudes.

### Validation Studies in Government and Industry

Numerous recent studies report on the use of achievement tests to predict training or job criteria in the armed forces, other governmental activities, or in industry. Sisson (35) reviewed development of Army and Navy personnel procedures from their origins in World War I thru World War II up to the present time. The *Army General Classification Test* and the *Navy General Classification Test* were described and validation data were presented. The development and validation of numerous other aptitude and achievement tests were described for such diverse areas as gunners' mates, radar operators, torpedomen, automotive mechanics, aircraft mechanics, radio mechanics, cooks, clerks, and machinists. The staff of the Personnel Research Section (37) described the development and validation of the currently utilized *Army Enlisted Classification Battery*. This battery consists of 10 tests which are processed to yield 10 composite scores for aptitude areas. As in most such batteries, several of the tests are essentially achievement tests. Intercorrelational and validity data are reported. Gragg and Gordon (20) reported the results of 66 validity studies on the currently utilized *Airman Classification Test Battery* in the Air Force. The tests, composite scores (aptitude indices), and years of education were correlated with final grades in the technical training schools.

Flanagan (12) briefly traced the development of aviation psychology to the present time and summarized the results of the World War II Army Air Force Aviation Psychology Program. A chart was presented showing validity of the pilot stanine for predicting success in primary pilot training. Numerous other validity studies were carried out for other aircrew positions as well as for private pilots and air-transport pilots. The extensive joint Air Force-Navy project on validation of the Air Force pilot tests with naval air cadets was described. He reported also extensive World War II work on the development of proficiency measures for instructors and aircrew with particular emphasis on objective flight checks for pilots. Dailey and Gragg (7) carried out extensive studies of the *Air Force Aviation Cadet Classification Battery* leading to its postwar revision. The validity of the battery for training success was found to be as high as in World War II despite important changes in both the training population and the nature of pilot training. It was found that the battery predicted elimination for flying deficiency much better than it predicted other categories of elimination, such as motivational elimination. Tupes and Cox (36) found that a combination of pilot information test (general information), a biographical inventory, and an attitude questionnaire yielded a multiple correlation of .61 with a criterion of motivational elimination in basic pilot training where the validity of the pilot stanine for the same sample and criterion was .34.

Zachert and Levine (39) found that years of education add little to the validity of the *Airman Classification Test Battery*. This battery included several tests that are essentially achievement tests. Littleton (28) found tests in arithmetic and blueprint reading to be valid for predicting instructor ratings in an auto trade course. Ghiselli and Brown (18) summarized a number of previously published validity studies for auto mechanics. They computed weighted-mean-validity coefficients and found the tests on arithmetic and mechanical principles to be among the most valid tests. Owens (29) conducted a validation study for the prediction of school grades in veterinary medicine. Highest validities were obtained for four new tests in chemistry achievement, zoology achievement, paragraph comprehension, and verbal memory. Lauer and Michael (27) described a new optometric test which included subjectmatter achievement sections in general culture and biology. DuBois (9) discussed the use of achievement and proficiency tests in civil-service-type examinations for purposes of selection. He concluded that achievement and aptitude tests are often interchangeable and recommended procedures for developing and using such tests.

### Factor Analyses of Achievement and Proficiency Tests

Several previously mentioned studies have suggested considerable overlap between the areas of achievement and aptitude tests. A number of studies have explicitly investigated this problem by means of factor analyses of combined matrices of achievement and aptitude tests and occasionally have included achievement and school grade criteria. Out of this work have come many intriguing insights into the nature of "aptitude" and "achievement" as measured by psychological tests. A greater understanding of the nature of many school and other criterion measures has also been accomplished. Much more work of this nature remains to be done, and work in this area should be encouraged. French (14) summarized the results of 64 factor analyses of aptitude and achievement tests previously published and described the 59 factors isolated. A number of these factors were defined by tests that were explicitly achievement tests. A number of such tests also had sizable saturations with factors normally regarded as aptitude factors. An attempt was made to differentiate between genetic and experimental factors. Fruchter (16) factored a matrix which included the parts of the *Army General Classification Test*, the *Airman Classification Test Battery*, the *Differential Aptitude Tests*, the *Gray-Votaw General Achievement Tests* (elementary science, social studies, knowledge of literature, choice of words, reading, and arithmetic), the *Iowa High School Content Examination*, and the *Otis Quick-Scoring Mental Ability Test*. He found several sections of the Gray-Votaw battery to have substantially the same factor content as similar subtests in the *Army General Classification Test*, the *Airman Battery*, and the *Differential Aptitude Tests*. The only new factor introduced by inclusion of the educational achievement batteries appeared to be a grammar factor. Doppelt and Wesman (8)

correlated the *Differential Aptitude Tests* with various educational achievement measures and found them to be highly correlated.

Various studies have obtained interesting results by incorporating criterion measures in the matrix to be factored. Bryant and Zachert (3) factored matrices of Airmen classification tests and Air Force technical school grades for clerk-typists and radar mechanics. Verbal, numerical, mechanical experience, academic information, visualization, perceptual speed, and general biographical background factors were isolated. Clerk-typist grades were found to be most heavily saturated with the verbal and numerical factors, while radar mechanic grades were more heavily saturated with the numerical and visualization factors. Comrey (5) factored a matrix of the tests in the *Air Force Aviation Cadet Classification Battery*, plus eight achievement grades at the Military Academy at West Point. He isolated the usual factors for that battery plus a new factor, which he labeled the "halo" factor. The academic measures vary considerably in factor content. French and others (15) did a factor study of 23 aptitude and achievement tests and 14 course grades at the United States Coast Guard Academy. Several previously identified factors were isolated plus a "Grade Aptitude" and an "Entrance Scores" factor, produced by the method of assigning entrance grades. Many "aptitude" and "achievement" tests in the battery showed considerable overlap in factor content. In a somewhat similar study, French (13) intercorrelated a number of aptitude and achievement tests plus 16 course grades for samples of students in the United States Coast Guard Academy and the Boston University General College. Without performing a factor analysis, it was found possible by examination of the clusterings of the intercorrelations to derive useful insights into the relationships between tests and specific grades and grade areas.

### Methodology for Proficiency Test Development and Evaluation

In recent years there has been a welcome trend toward a greater emphasis upon theoretical and experimental approaches to the problem of improving criteria for the validation of both aptitude and achievement tests. It has been recognized that the full development and fruition of the testing field depends upon advances in this area of proficiency measurement and criterion development. Gulliksen (22) recommended assessing achievement tests more systematically in terms of the concept of intrinsic validity. He suggested particularly the application of factor analysis to judgments of experts regarding test content and a more intensive use of pretraining and posttraining administration of tests. In a later statement, Gulliksen (21) suggested relating achievement tests to aptitude batteries and also factoring matrices of aptitude tests and criterion variables. He reported navy studies where the validity appeared to be too high for verbal tests and too low for mechanical tests for gunners' mates and torpedomen. Improvement of the proficiency measures in the two schools later reversed

this validity pattern. He also recommended validation of training achievement tests against later relevant measures of job success. Gorham (19) conducted a study of the selection of proficiency test items by means of internal consistency analysis as compared with the difference in item performance for groups before and after army basic recruit training. He recommended the latter method as being preferable. Brokaw (2) carried out an empirical test of formulas to estimate the effect that shortening tests in a battery of predictive tests has upon their prediction of a training criterion. His results verified the accuracy of the formulas, and indicated that cutting each test in half would reduce the multiple validity for an air force technical training school only negligibly. Several of his predictive tests were essentially achievement tests. Hausman, Begley, and Parris (23) developed and evaluated an orally administered achievement test in aircraft maintenance. It was demonstrated that the new test had less verbal-factor variance than an equivalent written test and also had good validity for supervisor ratings and showed good "customer acceptability." Cureton (6) has given a comprehensive summary of much current work and thinking on the problems of test validation. His presentation emphasized the vital importance of criterion logic and analysis in the validation process and the complexity of most current approaches to the problem of defining and measuring the behaviors to be predicted. He also discussed several statistical problems involved in criterion analysis and in validation. Ryans and Frederiksen (34) discussed the area of development and evaluation of performance tests of educational achievement. This area was defined broadly to include all types of nonwritten tests of the results of instruction. Numerous examples of such tests were described and suggestions given for their optimal use. Theoretical aspects of such test development and evaluation were covered comprehensively, and a detailed and useful outline of a procedure for the development of performance tests was presented.

### Achievement Tests for Professional Fields

Baier, Harmon, and McAdoo (1) developed and validated a *Statistics Test* and demonstrated successful use of it in training the staff of the Personnel Research Section of the Army Adjutant General's Office. Jouno (25) described the development and use of the *Federal Junior Professional Assistant Examination*. In this examination competitors in all options took an aptitude-information test of general verbal abilities and quantitative abilities and also took subjectmatter tests in their option. Findley (11) developed novel types of tests to measure ability to solve realistic field situation problems at the Air Force Air University.

### Bibliography

1. BAIER, DONALD E.; HARMON, HARRY H.; and MCADOO, HAROLD L. "Can Personnel Researchers Test and Train Themselves in Statistics?" *Educational and Psychological Measurement* 12: 267-74; Summer 1952.
2. BROKAW, LELAND D. "Comparative Validities of 'Short' Versus 'Long' Tests." *Journal of Applied Psychology* 35: 325-30; October 1951.

3. BRYANT, NORMAN D., and ZACHERT, VIRGINIA. *Factor Analyses of the Airman Classification Battery with Criteria for Clerk-Typist and Radar Mechanic Technical Schools*. Research Bulletin 51-22. Lackland Air Force Base, San Antonio, Texas; Human Resources Research Center, 1951. 20 p.
4. CHAUSOW, HYMEN M. "The G. E. D. and the Social Studies." *Junior College Journal* 22: 450-56; April 1952.
5. COMREY, ANDREW L. "Factorial Study of Achievement in West Point Courses." *Educational and Psychological Measurement* 9: 193-209; Summer 1949.
6. CURETON, EDWARD E. "Validity." *Educational Measurement*. (Edited by Everett F. Lindquist.) Washington, D. C.: American Council on Education, 1951. Chapter 16, p. 621-94.
7. DAILEY, JOHN T., and GRAGG, DONALD B. *Postwar Research on the Classification of Aircrew*. Research Bulletin 49-52. Lackland Air Force Base, San Antonio, Texas; Human Resources Research Center, 1949. 64 p.
8. DOPPELT, JEROME E., and WESMAN, ALEXANDER G. "The Differential Aptitude Tests as Predictors of Achievement Test Scores." *Journal of Educational Psychology* 43: 210-17; April 1952.
9. DUBOIS, PHILIP M. "Achievement Tests in Personnel Selection." *American Journal of Public Health* 41: 567-75; May 1951.
10. EDUCATIONAL TESTING SERVICE. *Annual Report to the Board of Trustees 1950-51*. Princeton, N. J.: 1952. 105 p.
11. FINDLEY, WARREN G. "Transferring Field Situations to Test Exercises at the Air University." *Proceedings of the 1948 Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1949. p. 25-27.
12. FLANAGAN, JOHN C. "Aviation." *Handbook of Applied Psychology*. (Edited by Douglas H. Fryer and Edwin R. Henry.) New York: Rinehart and Co., 1950. Chapter 7, p. 341-48.
13. FRENCH, JOHN W. "An Analysis of Course Grades." *Educational and Psychological Measurement* 11: 280-94; Summer 1951.
14. FRENCH, JOHN W. *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors*. Psychometric Monographs, No. 5. Chicago: University of Chicago Press, 1951. 278 p.
15. FRENCH, JOHN W. and OTHERS. "Factor Analyses of Aptitude and Achievement Entrance Tests and Course Grades at the U. S. Coast Guard Academy." *Journal of Educational Psychology* 43: 65-80; February 1952.
16. FRUCHTER, BENJAMIN. "Orthogonal and Oblique Solutions of a Battery of Aptitude, Achievement, and Background Variables." *Educational and Psychological Measurement* 12: 20-38; Spring 1952.
17. FUESS, CLAUDE M. *The College Board: Its First Fifty Years*. New York: Columbia University Press, 1950. 222 p.
18. GHISELLI, EDWIN E., and BROWN, CLARENCE W. "Validity of Tests for Auto Mechanics." *Journal of Applied Psychology* 35: 23-24; February 1951.
19. GORHAM, WILLIAM A. *A Comparative Study of Two Internal Criteria for the Selection of Valid Proficiency Test Items*. Cleveland, Ohio: Western Reserve University, 1950. (Master's thesis)
20. GRAGG, DONALD B., and GORDON, MARY A. *Validity of the Airman Classification Battery AC-1*. Research Bulletin 50-3. Lackland Air Force Base, San Antonio, Texas; Human Resources Research Center, 1950. 266 p.
21. GULLIKSEN, HAROLD. "Criteria for the Evaluation of Achievement Tests: From the Point of View of Their External Statistical Relationships." *Proceedings of the 1950 Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1951. p. 100-103.
22. GULLIKSEN, HAROLD. "Intrinsic Validity." *American Psychologist* 5: 511-17; October 1950.
23. HAUSMAN, HOWARD J.; BEGLEY, JOSEPH J.; and PARRIS, HOWARD L. "Oral Examinations for Proficiency Testing." (Abstract) *American Psychologist* 5: 362-63; July 1950.
24. JONES, EDWARD S. "Some Results of Requiring the Graduate Record Examination of All Seniors." *Educational Record* 33: 105-10; January 1952.
25. JOUNO, RANDOLPH J. "The Federal Junior Professional Assistant Examination." *Occupations* 28: 361-63; March 1950.
26. LANNHOLM, GERALD V., and SCHRADER, WILLIAM B. *Predicting Graduate School Success*. Princeton, N. J.: Educational Testing Service, 1951. 50 p.

27. LAUER, ALVAH R., and MICHAEL, WILLIAM B. "Evaluation of an Optometric Test." *Educational and Psychological Measurement* 10: 685-92; Winter 1950.
28. LITTLETON, ISAAC T. "Prediction in Auto Trade Courses." *Journal of Applied Psychology* 36: 15-19; June 1952.
29. OWENS, WILLIAM A. "An Aptitude Test for Veterinary Medicine." *Journal of Applied Psychology* 34: 295-99; October 1950.
30. ROEBER, EDWARD C. "GED Tests as a Measure of College Aptitude." *Educational Research Bulletin* 29: 40-41; February 1950.
31. RYANS, DAVID G. "An Analysis of Teacher Examination Scores of College Seniors Who Expect to Become Teachers." (Abstract) *American Psychologist* 4: 288; July 1949.
32. RYANS, DAVID G. "National Teacher Examinations: Their Use in the Selection of Teachers." *Proceedings of the Thirty-Fifth Annual Schoolmen's Week*. Philadelphia: University of Pennsylvania, 1948. p. 271-81.
33. RYANS, DAVID G. "The Results of Internal Consistency and External Validation Procedures Applied in the Analysis of Test Items Measuring Professional Information." *Educational and Psychological Measurement* 11: 549-60; Winter 1951.
34. RYANS, DAVID G., and FREDERIKSEN, NORMAN. "Performance Tests of Educational Achievement." *Educational Measurement*. (Edited by Everett F. Lindquist.) Washington, D. C.: American Council on Education, 1951. Chapter 12, p. 455-94.
35. SISSON, E. DONALD. "Military Personnel Management." *Handbook of Applied Psychology*. (Edited by Douglas H. Fryer and Edwin R. Henry.) New York: Rinehart and Co., 1950. Chapter 5, p. 236-47.
36. TUPES, ERNEST C., and COX, JOHN A. *Prediction of Elimination from Basic Pilot Training for Reasons Other Than Flying Deficiency*. Research Bulletin 51-1. Lackland Air Force Base, San Antonio, Texas: Human Resources Research Center, 1951. 25 p.
37. U. S. ARMY, ADJUTANT GENERAL'S OFFICE. *Development of Aptitude Areas for Classification of Enlisted Personnel in the Army*. Personnel Research Section Report No. 808. Washington, D. C.: Adjutant General's Office, 1949. 37 p.
38. WARDLAW, H. PAT. "Use and Value of GED Tests for College Entrance of Veterans in the Armed Forces." *North Central Association Quarterly* 26: 295-301; January 1952.
39. ZACHERT, VIRGINIA, and LEVINE, ABRAHAM S. "Education and Prediction of Military School Success." *Journal of Applied Psychology* 36: 266-68; August 1952.

## Index to Volume XXIII, No. 1

Page citations are made to single pages. These are often the beginning of a chapter, section, or running discussion relating to the topic.

- Achievement, and motivation, 93; and personality, 93
- Achievement tests, 75, 102; in evaluation of school methods and policies, 90; factor analysis of, 105; norms, 88; and prediction, 92; reliability, 87; scaling methods, 89; validity, 87
- Aptitude tests, 33; accounting, 37; administrative, 37; art, 41; clerical, 41; dental, 36; engineering, 35; general aptitude, 38; legal, 35; mechanical ability, 41; medical, 34; music, 41
- Circumvention, of intent of inventories, 60
- Differential prediction, 39
- Evaluation, practices, 6; of proficiency tests, 106; of school methods and policies, 90
- Factor analysis, 57, 105
- General educational development, 102
- Intelligence, and environmental factors, 14; and physical factors, 14
- Intelligence tests, 11; brief, 20; for children, 18; culture-free, 19
- Interests, inventories, 56; measurement of, 56
- Inventories, applications, 61; interest, 56; personality, 56
- Item selection, 45
- Norms, achievement tests, 88; personality and interest inventories, 59; projective technics, 75
- Personality, and achievement, 93; inventories, 56; nonprojective tests of, 56; tests, 70
- Prediction, and achievement tests, 92; differential, 39; scholastic achievement, 40
- Projective technics, 70; applications, 76; norms, 75; reliability, 70; validity, 70
- Reliability, 46; achievement tests, 87; projective technics, 70; personality and interest inventories, 59
- Scaling methods, achievement tests, 89
- Scoring, 47
- Selection, of items, 45; of tests, 44
- Selective Service tests, 33
- Test theory, 48, 62
- Testing programs, 5
- Tests, accounting aptitude, 37; achievement, 85, 102; administrative aptitude, 37; aptitude, 33; art, 41; bibliographic sources, 5; clerical, 41; culture-free, 19; dental aptitude, 36; development of, 48; discriminating power, 63; engineering aptitude, 35; GED, 102; general aptitude, 38; of general mental ability, 11; intelligence, 11; interest, 56; legal aptitude, 36; measurement of, 56; mechanical ability, 41; medical aptitude, 34; music, 41; nonprojective, 56; personality, 56, 70; proficiency, 102; scoring, 47; selection of, 44
- Textbooks, educational and psychological measurement, 7; on mental testing, 11
- Validation studies, in government and industry, 104
- Validity, 104; achievement tests, 87; of aptitude tests, 43; and the criterion, 43; of inventories, 61; projective technics, 70